Ingenieurfakultät Bau Geo Umwelt
Signalverarbeitung in der Erdbeobachtung

Multi-sensor Data Fusion for Multi- and Hyperspectral Resolution Enhancement Based on Sparse Representations

Claas Hendrik Grohnfeldt

Vollständiger Abdruck der von der Ingenieurfakultät Bau Geo Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr.-Ing.)

genehmigten Dissertation.

Vorsitzende:
Prof. Dr.-Ing. habil. Liqiu Meng

Prüfende der Dissertation:
1. Prof. Dr.-Ing. habil. Xiaoxiang Zhu
2. Prof. Dr.-Ing. habil. Richard Bamler
3. Assistant Prof. Dr. Naoto Yokoya
   University of Tokyo, Japan

Die Dissertation wurde am 29.06.2017 bei der Technischen Universität München eingereicht und durch die Ingenieurfakultät Bau Geo Umwelt am 31.08.2017 angenommen.

Abstract

Our world is subject to intense observation. Offshore zones, arctic areas, deserts, rainforests, rivers, agricultural land and cities are continuously monitored from space. There is no surface on Earth that goes unobserved by a remote sensing satellite. This transparency facilitates our modern life, offering seemingly inexhaustible opportunities. Yet, there are limits.

Physical, technological and financial factors limit the development of sensors of ever-increasing accuracy. Trade-offs must be made. Optical imaging sensors compromise on either detailed spectral information, allowing for the discrimination of materials, or high spatial resolution, which elucidates the geometry of the scene.

Multi-sensor data fusion mitigates these limitations. The idea is to combine complementary data of different spectral and spatial characteristics to construct products representative of an "ideal" synthetic sensor. This is an inherently ill-posed problem. The difficulty arises in maximizing both spatial and spectral resolutions, due to the number of degrees of freedom.

This dissertation presents sophisticated solutions to two of the most demanding multi-sensor data fusion problems in remote sensing, namely pan-sharpening and the fusion of hyperspectral and multispectral data. Two algorithms are thoroughly designed and tailored to the respective problems, i.e., Jointly Sparse Fusion of Images (J-SparseFI) for pan-sharpening and Jointly Sparse Fusion of Hyperspectral and Multispectral Imagery (J-SparseFI-HM) for the fusion of hyperspectral and multispectral data.

These severely ill-posed problems are handled by incorporating additional information, to reduce the inherent degrees of freedom and produce fusion products that better approximate the ground truth. The physical properties of the synthesized sensor are accounted for by a sensor observation model. Prior knowledge about image patches featuring sparse representations of suitable dictionaries is exploited to reduce the overall fusion problem to a large number of small, $l_{1,2}$-regularized convex optimization problems. Prior information about the mutual correlation of adjacent multi- and hyperspectral channels is extensively used to robustify estimation accuracies. Both techniques leverage distributed compressive sensing theory, restricting the solution of an underdetermined system by considering a jointly sparse ensemble of signals.

The J-SparseFI and J-SparseFI-HM algorithms are specifically tailored to their respective fusion problems. In J-SparseFI, a decision-based spectral and channel mutual correlation

analysis is conducted to find suitable groups of multispectral channels and corresponding high-resolution source data for dictionary learning. J-SparseFI-HM incorporates a spatially adaptive Correlation-based HyperSpectral Grouping (CorHySpeG) concept and the data for dictionary learning is found via non-negative least squares regression. A novel alternating local–non-local–global optimization procedure is designed for J-SparseFI-HM, which enhances spatial consistency across individually processed patches and spectral consistency across group-wise processed hyperspectral channels.

Highly parallel stand-alone software solutions are optimized for operation on the SuperMUC supercomputer of the Leibniz Supercomputing Center. Detailed parameter descriptions, sensitivity analyses and default setting recommendations are provided. The quality of the data fusion products is assessed and compared to the state-of-the-art in pan-sharpening and hyperspectral-multispectral image fusion, for a large variety of test scenarios and sensor combinations.

This thesis comprehensively demonstrates that J-SparseFI and J-SparseFI-HM are new quality benchmarks for multiresolution multi-sensor data fusion.

Kurzfassung

Unsere Welt wird intensiv beobachtet. Tiefseegebiete, arktische Gebiete, Wüsten, Regenwälder, Flüsse, Agrarlandschaften und Städte werden kontinuierlich überwacht. Es gibt keine Oberfläche auf der Erde, die nicht von Satelliten erkundet wurde. Diese Transparenz erleichtert unser modernes Leben und schafft scheinbar unersättlich viele Möglichkeiten. Jedoch gibt es Einschränkungen.

Physikalische, technologische und finanzielle Faktoren limitieren die Entwicklung von Sensoren stetig höherer Genauigkeit und erzwingen Kompromisse. Optische Abbildungssensoren verfügen entweder über hohe spektrale Kapazitäten, mit Hilfe derer Materialien unterschieden werden können, oder hohe räumliche Auflösung, welche die Geometrie der Szene zu erkennen gibt.

Multi-Sensor-Datenfusion löst diese Einschränkungen auf. Dabei werden komplementäre Daten unterschiedlicher spektraler und räumlicher Eigenschaften zu einem Datenprodukt eines "idealen" synthetischen Sensors kombiniert. Dieses Problem ist in natürlicher Weise unterbestimmt. Die Anzahl der Freiheitsgrade – und somit die Schwierigkeit – nimmt mit wachsender Diskrepanz zwischen den jeweiligen Auflösungen rapide zu.

Diese Dissertation präsentiert technisch ausgereifte Lösungen für zwei der anspruchsvollsten und relevantesten Multi-Sensor-Datenfusionsprobleme in der Fernerkundung: Pan-Schärfung und Hyper- und Multispektrale Datenfusion. Entsprechende Algorithmen mit den Namen "Jointly Sparse Fusion of Images" (J-SparseFI) und "Jointly Sparse Fusion of Hyperspectral and Multispectral Imagery" (J-SparseFI-HM) sind detailliert entwickelt und sorgfältig auf die jeweiligen Probleme zugeschnitten.

Diese hochgradig unterbestimmten Probleme sind behandelt, indem eine Vielzahl von Informationen berücksichtigt sind, welche die Anzahl der Freiheitsgrade verringern und die Fusionsprodukte die exakten Bodeninformationen genauer approximieren lassen. Die physikalischen Eigenschaften des synthetisierten Sensors sind durch ein Sensor-Observations-Modell berücksichtigt. Vorwissen über die dünn-besetzte Darstellbarkeit von Bildausschnitten in geeigneten Dictionaries ist ausgenutzt, um das Gesamtproblem in viele kleine $l_{1,2}$-regularisierte, konvexe Optimierungsprobleme aufzuteilen. Korrelationen zwischen benachbarten Multi- und Hyperspektralkanälen ist genutzt, um die Koeffizientenschätzung zu verbessern. Beide Algorithmen setzen Techniken der Distributed Compressive

Sensing Theorie ein, die den Lösungsraum unterbestimmte Systeme einschränken, indem sie gruppierte Dünn-besetztheit von Signalen fördern.

Der J-SparseFI Algorithmus beinhaltet eine entscheidungsbasierte Spectral- und Korrelationsanalyse zur Identifikation geeigneter Kanalgruppen und zugehöriger hochaufgelöster Quelldaten für die Erzeugung der Dictionaries. Für J-SparseFI-HM wurde ein adaptives, korrelationsbasiertes Hyperspektralgruppierungsverfahren (CorHySpeG) entwickelt und die Quelldaten für die Erzeugung der Dictionaries wird mittels nicht-negativer Kleinste-Quadrate-Regression berechnet. Ein local–non-local–global Optimierungsverfahren wurde für J-SparseFI-HM entwickelt, welches die Konsistenz über individuell prozessierte Bildausschnitte und gruppenweise prozessierte Hyperspektralkanäle erhöht.

Hochparallele Softwareprodukte wurden entwickelt und für Datenprozessierung auf dem SuperMUC Supercomputer des Leipniz Rechenzentrums optimiert. Detaillierte Parameterbeschreibungen, Sensitivitätsanalysen und Hinweise auf geeignete Einstellungen sind bereitgestellt. Die Qualität der Datenfusionsprodukte ist für eine Vielzahl von Sensorkombinationen und Testszenarien evaluiert und mit dem Stand der Technik verglichen. In dieser Arbeit ist umfassend demonstriert, dass J-SparseFI und J-SparseFI-HM neue Maßstäbe im Bereich der mehrfach aufgelösten Multi-Sensor-Datenfusion setzen.

Contents

1

Introduction

1.1 Motivation and Objectives

Due to physical, technological and financial limitations, trade-offs have to be made during the design phase of a sensor in order to ensure it meets the design requirements and mission objectives. In the case of optical imaging sensors, the most prominent trade-off is made between spatial resolution, spectral fidelity and signal-to-noise ratio. Additional constraints are introduced for satellite-based optical instruments, used in Earth observation, such as the long distance between the carrier platform and the observed area as well as atmospheric absorption and scattering processes, which influence the measured signals. Hence, optical Earth observation systems are constrained to provide *either* detailed spectral information *or* high spatial resolution; the finer the spectral bandwidth, the lower the spatial resolution and vice versa.

The highest spatial resolution is obtained with sensors that feature a single broadband channel. Those sensors are called *panchromatic* sensors. They are typically sensitive to energy over a large portion of the visible wavelength range, which means that they collectively map energy measured in different wavelengths into a single channel which is usually visualized as grayscale imagery. The highest spatial resolution currently achievable with commercial satellite-based panchromatic sensors is around 0.3–0.5 m ground resolved distance.

In order to distinguish colors as they are seen by the human eye (so-called *true colors*), at least three channels are required, each of which measures incoming light in a relatively narrow region of the spectrum, containing red (ca. 620–750 nm), green (ca. 495–570 nm) or blue (ca. 450–495 nm) wavelengths, respectively. Since each of those so-called *multispectral* channels individually measure less incoming light in a given time interval and instantaneous field of view than a broadband panchromatic channel of the same spatial resolution, the signal-to-noise ratio of the multispectral channels is lower. In order to keep the signal-to-noise ratio – and, therefore, the quality of the data – high, the instantaneous fields of view of the multispectral channels are broadened, so that more light, i.e. signal, can reach the detectors. Consequently, the spatial resolution of multispectral imagers is generally lower than those of panchromatic sensors; typically between 1.5–10 m ground resolved distance.

If a sensor comprises multiple channels, each pixel features a *spectral profile*, which is a

wavelength-dependent sequence of numbers each of which quantifies the amount of light measured in a specific channel. Such spectral information is used by the human eye and data analysis algorithms to associate pixels with types of material. For instance, pixels, whose spectral profiles have a dominant entry in a channel that is sensitive to green light, will appear green and are likely to be associated with vegetation. This is because, vegetation *reflects* some of the solar irradiance in this wavelength range while being more *absorptive* to red and blue light. Following this principle, each material has a unique continuous *spectral signature* indicating its relative reflectivity as a function of wavelength. Hence, more numerous, narrow spectral channels result in higher fidelity spectral profiles and improve the sensor's capacity to discriminate between materials. Considering that there are only three color-sensitive cones in the retina of the human eye, corresponding roughly to red, green, and blue, computer-aided analysis and processing of highly resolved spectral data enables the identification and discrimination of materials far beyond the capacity of humans. Sensors constructed to provide such data are called *hyperspectral* sensors and are characterized by tens to hundreds of narrow spectral channels in the visible to near infrared or even shortwave infrared wavelength range, i.e., 400–2500 nm. One of the most advanced satellite-based hyperspectral sensors is the imaging spectrometer of the upcoming Environmental Mapping and Analysis Program (EnMAP) – a German hyperspectral satellite mission that aims at monitoring and characterising the Earth's environment on a global scale. Despite the state-of-the-art sensor technology, the aforementioned trade-off between spectral and spatial resolution is unavoidable. In particular, the spatial resolution of EnMAP is expected to be 30 m ground resolved distance. The challenge implicated by such a low spatial resolution is that multiple materials in a given scene are mixed in the same pixels. Especially in heterogeneous regions, such as urban areas, that are characterized by both large varieties of materials and high levels of spatial details, there is a high demand for high-spatial-high-spectral resolution data. Since there is no spaceborne sensor that can provide such precious data at present or in the near future, scientists have recently been exploring signal processing techniques to overcome these limitations. As is explained below, multi-sensor data fusion is an excellent approach in this regard.

Multi-sensor data fusion aims at combining data acquired by complementary sensors to obtain a more compact and complete representation of the information contained in the individual input data. Considering the complementary spatial and spectral characteristics of panchromatic, multispectral and hyperspectral sensors, they present an ideal basis for multi-sensor data fusion. This knowledge, combined with the availability of satellite systems that carry both a panchromatic and a lower-resolution multispectral sensor onboard, has led to the development of numerous multispectral-panchromatic data fusion methods in the past. This particular fusion problem is commonly referred to as *pan-sharpening*. Pan-sharpening aims to produce data that would be acquired by a sensor featuring the spectral characteristics of the multispectral sensor and the spatial characteristics of the higher-resolution panchromatic sensor. On the other hand, sharpening of hyperspectral imagery via data fusion has been a less actively pursued topic until very recently. This because of the limited availability of hyperspectral satellite imagery in the past and the comparatively small size of the community working with hyperspectral data. However, multiple upcoming hyperspectral satellite missions are currently leading to steady growth of the hyperspectral remote sensing community, increasing interest in finding new applications based on spaceborne hyperspectral data, and demands for algorithms to enhance the quality and spatial resolution of hyperspectral data. Initial attempts to fuse hyperspectral with higher-resolution multispectral or panchromatic data have been proposed in the recent past, however, there is considerable room for improvement.

The reason why multiresolution data fusion problems remain the focus of active research, is because the problems have no universal solutions. This is obvious when considering the simple fact that, if the spatial resolution ratio between the two input images is at least equal to two, which is always the case, the dimensions of both input images are, collectively, lower than the dimension of the fusion product. Hence, the problem of finding a higher-dimensional fusion product is inherently ill-posed. The larger the spatial resolution ratio between both input images, the larger the number of degrees of freedom. Similarly, the larger the difference between the numbers of spectral channels, the larger the number of degrees of freedom.

In order to find a solution that is as close to the unknown "true" high-resolution multi- or hyperspectral surface information as possible, it is advantageous to solve the fusion task based on a mathematical model that incorporates as much information about the original sensing systems, input data, synthesized sensor and fusion product as is available, to describe the inter-sensor relationship and link the input to the output data. As will be substantiated by comparative tests presented in Chapters 4 and 5, fusion approaches not utilizing sensor information about spectral and spatial sensitivity distributions, for instance, generally produce data of lower quality than those doing so. Such a model adds physical meaning to the fusion product and restricts the solution space that comprises all potential fusion products. Further assumptions need to be made to render the fusion task well-posed – assumption that can be used, for instance, to formulate the fusion task as a constrained or regularized underdetermined inverse problem.

There are a pool of theories that are being developed, primarily by the applied mathematical community, for solving different types of underdetermined linear systems. Among those, one of the most actively developed theories recently concerns the recovery of *sparse* signals. In numerous engineering fields involving signal processing, sparsity has been found to be a powerful prior for the robust and accurate identification of signals, especially in the presence of noise on the measurements. The number and diversity of applications utilizing the prior knowledge of signals featuring a sparse representation in the underlying system matrix or dictionary are increasing rapidly. Accordingly, initial attempts to apply the concept of sparsity to multi-sensor data fusion were reported very recently, with promising results achieved. This dissertation follows up on this development and further pushes the limits in what will be elaborately demonstrated to be one of the most reliable, robust and high quality approaches to enhancing the spatial resolution of multi- and hyperspectral data today. In particular, sophisticated sparse representation-based solutions to the general pan-sharpening and hyperspectral-multispectral data fusion problems are designed and presented separately, each being thoroughly tailored to the respective fusion task.

1.2 Structure of the Thesis

Chapter 2 introduces the fundamentals of optical satellite remote sensing, multiresolution data fusion and sparse signal recovery. A review of related literature and a summary of the contributions of this dissertation are provided in Chapter 3. Chapters 4 and 5 present the data fusion algorithms that have been designed in the scope of the research towards this dissertation to solve the pan-sharpening and the hyperspectral-multispectral data fusion problems, respectively. Each of those two chapters introduce the corresponding methodology and provide thorough parameter analyses and comprehensive fusion quality assessments relative to the state of the art. Conclusions are drawn and an outlook on future work is given in Chapter 6.

2

Basics

This chapter introduces relevant fundamentals of satellite remote sensing optical instruments, multiresolution data fusion and sparse signal recovery. It lays the theoretical foundation that this work's contribution is based upon and defines the notation used in the remainder of this thesis.

2.1 Optical Satellite Remote Sensing

Remote sensing systems can be divided into various categories. One important discrimination can be made between active and passive systems. While all remote sensing systems are equipped with a receiver, which detects (i.e. "senses") incident energy, only active sensors additionally use a transmitting device that sends out signals whose reflections are measured by the receiver of the same system. That is, passive sensors measure energy originating at an external source while active systems carry their own energy source.

Moreover, different fields of application implicate different constraints and requirements on the remote sensing system subject to which the sensor is designed and characterised. Prominent fields of application include, but are not limited to, Earth observation, medical imaging, astronomy and photography.

In Earth observation, another aspect according to which remote sensing systems can be differentiated by is the type of platform carrying the sensor. The main categories here include satellite-based, airborne (i.e. airplane-based) and UAV-based sensors.

The by far strongest and most reliable energy source utilizable by passive Earth observing sensors is the Sun, which is why the majority of passive systems operate in the Sun's irradiative wavelength range. Systems measuring solar energy are additionally restricted to those portions of the spectrum that are not blocked by the atmosphere. Those constraints are further discussed below in Section 2.1.1.

The broader spectral region in which both the Sun is emissive and the atmosphere is transmissive is called *electro-optical and infrared* wavelength range, which is sometimes simply referred to as *optical* wavelength range. This thesis concentrates on passive space-based Earth

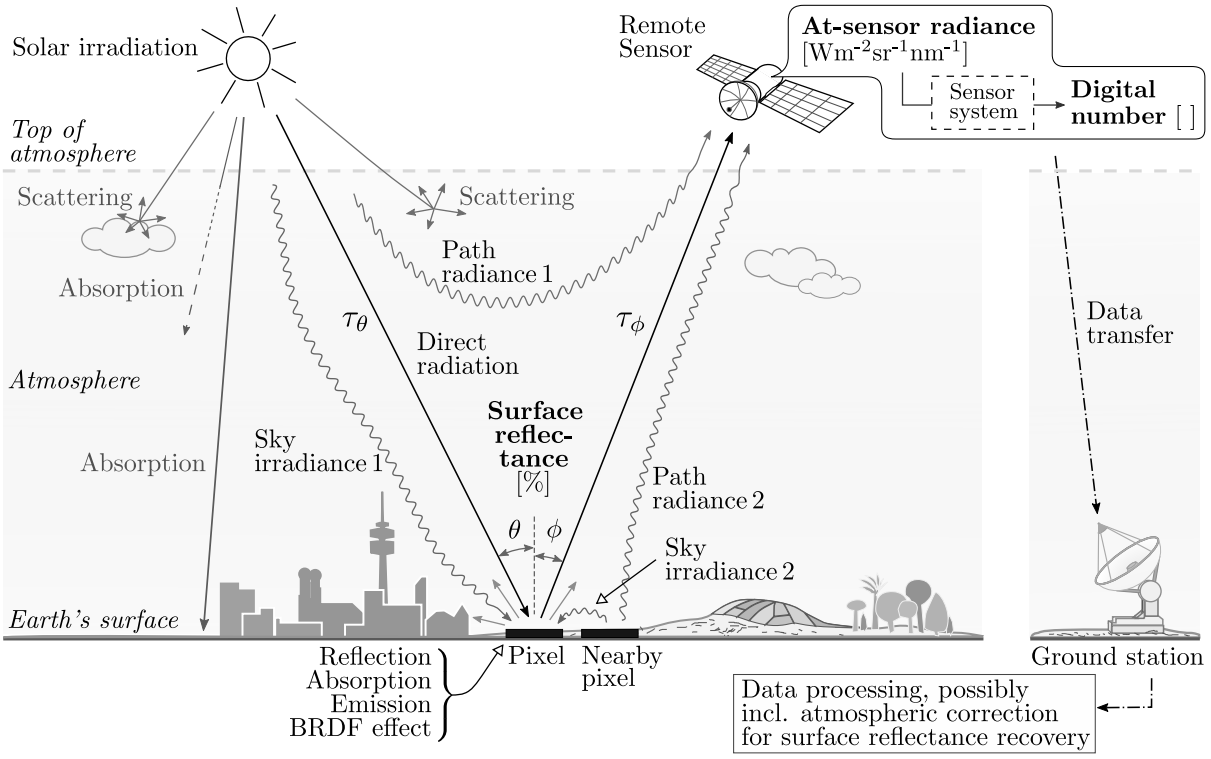


Fig. 1. Data acquisition process with a passive optical satellite remote sensing system and the effect of the atmosphere on the signal reaching the remote sensor. This measured signal is a superposition of various radiance components – mainly originating at the Sun – which are transferred through the atmosphere. In the atmosphere, the radiation is reflected by the observed area on the Earth's surface (straight line pointing at the satellite) or scattered into the field of view of the sensor (Path radiances 1 & 2). Atmospheric opacity to solar energy is mainly caused by absorption and scattering processes in the atmosphere. In addition to the direct radiation reaching the ground (straight line pointing down at the pixel), the observed surface area is illuminated by radiation reflected by nearby objects and solar energy redirected within the atmosphere (Sky irradiances 1 & 2). On the ground, radiation is partially absorbed, reflected or emitted[1]. The surface material dependent processes that relate the incident irradiance to the reflected radiance can be comprehensively described by a bidirectional reflectance distribution function (BRDF). Depending on the product processing level, optical remote sensing data is commonly available as at-sensor radiance, digital number (DN) or reflectance values. *(Graphics and caption contain cumulative information originating from Richards [2013, p. 34], Avbelj [2015, p. 18], Chuvieco and Huete [2009, p. 3], Schowengerdt [2006, p. 76] and Eismann [2012, pp. 146-148])*

observing remote sensing systems operating in the optical wavelength range. Although the developed theory and algorithms are principally applicable also to airborne, UAV-based and even medical and other imaging systems, corresponding applications are not explicitly discussed.

The remainder of this section summarizes some of the basic concepts of optical satellite remote sensing. Most of the terminology used in Chapters 4 and 5 is introduced here. Specifically, the subjects discussed include physical and imaging principles, sensor characteristics and resolutions, remote sensing data processing steps, data representations and signal processing aspects, and an overview of relevant Earth observation satellite missions.

2.1.1 Physical and Imaging Principles

Optical remote sensing data can be acquired of anything measurable from remote in the optical electromagnetic (EM) wavelength range. What exactly *is* measured and *how* it is measured are the subjects addressed in this section.

Fig. 1 depicts an overview of the data acquisition process and the influence of the atmosphere on the signal measured by a passive optical Earth observing satellite remote sensing

system. In order to avoid an extensive distinction of cases while explaining this mechanism, it is purposeful to temporarily ignore thermal scanners and at-night observations. Then, the main contribution to the radiation ultimately reaching the sensor originates at the Sun and is *reflected* by the observed area on the Earth's surface. Before arriving at the Earth's atmosphere, the solar energy radiates through space isotropically. Within the atmosphere, this energy is partially absorbed, transmitted, scattered and redirected. The relative amount of solar radiation directly reaching the ground in an incidence angle of θ is called *transmittance*, which is denoted by τ_θ. Similarly, there is an atmospheric transmittance, τ_ϕ, between the point of reflection and the sensor. Mainly due to scattering processes, the energy originally radiating in other directions is redirected by the atmosphere and falls onto the observed pixel (Sky irradiances 1 & 2) or into the sensor's field of view (Path irradiances 1 & 2). On the ground, incident radiation is partially absorbed, reflected or emitted [1]. The process relating the incident irradiance to the reflected radiance is described – for non-Lambertian surfaces – by the bidirectional reflectance distribution function (BRDF) [Nicodemus, 1965; Andrews et al., 2015], which can be measured for particular 3-D physical structures on the ground, such as different types of vegetation, and specific angles θ and ϕ. Since those parameters are partially unknown under general satellite operating conditions, the surface is commonly assumed Lambertian. In this case, the BRDF can be approximated using a diffuse spectral surface reflectance model [Schowengerdt, 2006, p. 53]. Currently, there is an operational satellite-based hyperspectral remote sensing mission by the name of CHRIS/PROBA, whose primary objective is the collection of BRDF data for a better understanding of spectral surface reflectance [Barnsley et al., 2004].

Depending on the processing level, optical remote sensing data is typically available in one of the following formats:

- ◇ Digital number (DN) [ ]: A unitless quantity representing the raw data as measured by the sensor.
- ◇ At-sensor radiance, L, [$\mathrm{Wm^{-2}sr^{-1}nm^{-1}}$]: The total radiance reaching the sensor, i.e. the signal that has not yet travelled through the various instrument components of the sensing system.
- ◇ Surface reflectance, ρ, [%]: The unitless ratio, defined as the reflected radiance normalized by the incident irradiance [Bachmann, 2007]. The reflectance rate of the observed scene can be set in simple relationship to the scene's surface absorption rate, α, and surface transmittance rate, τ_{surface}, as follows [see e.g. Bieniarz, 2015]:

$$\rho = 1 - \alpha - \tau_{\mathrm{surface}} \tag{1}$$

Commencing with the digital number, the at-sensor radiance can be reconstructed by accounting for the known sensor system properties that are partially measured during the calibration phase, as described further below in this section. Under the assumption of a Lambertian surface model, this quantity is given by [Schowengerdt, 2006, pp. 52-55]

$$L = \rho \frac{\tau_\phi}{\pi} E_g + L_p + L_e\,, \tag{2}$$

where ρ and τ_ϕ are the above-defined surface reflectance and the atmospheric transmittance, and

- ◇ L_p is the path radiance, constituted by the path radiances 1 & 2 as shown in Fig. 1;
- ◇ L_e is the emissive radiance as defined in Schowengerdt [2006, p. 68];

[1] Emission occurs mainly in the thermal portion of the spectrum as will be described later in this section.

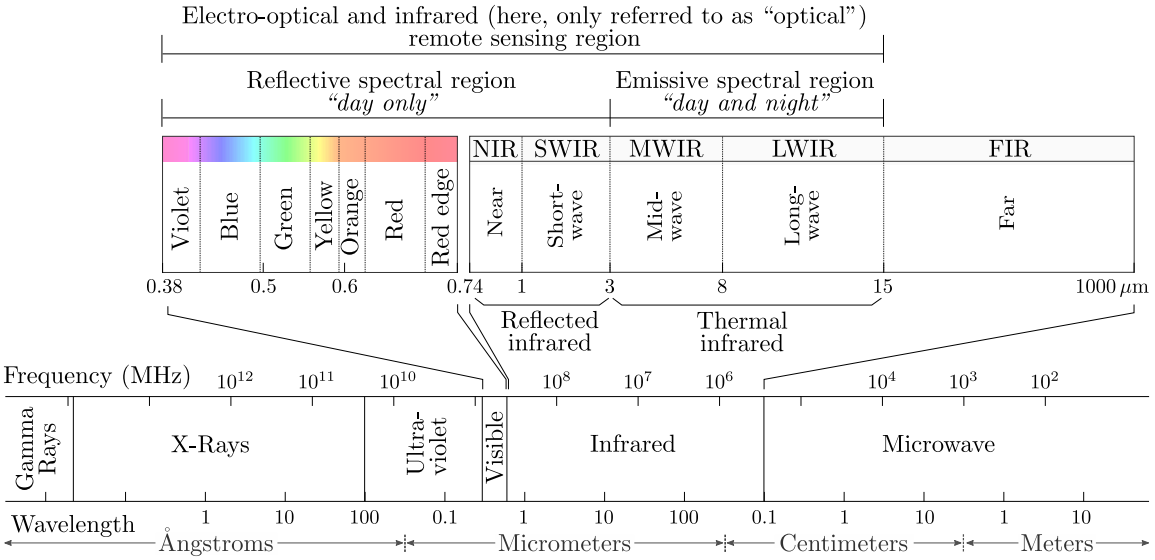


Fig. 2. Electromagnetic spectrum from gamma rays to microwaves with the visible and infrared regions indicated. The wavelength range from just below 0.4 μm to 15 μm is where electro-optical and infrared remote sensing systems operate. Signals measured in the spectral portion from around 0.4 μm (visible spectrum) to 3 μm (shortwave infrared spectrum) correspond to reflected solar radiation, which constraints the data acquisition planning to daylight. Signals in the midwave and longwave infrared range, on the other hand, correspond to thermal energy. This energy is emitted independently of solar irradiance, i.e. both during the day and at night. *(Graphics and caption contain cumulative information originating from Chuvieco and Huete [2009, p. 24], Alparone et al. [2015, p. 30] and Eismann [2012, p. 3])*

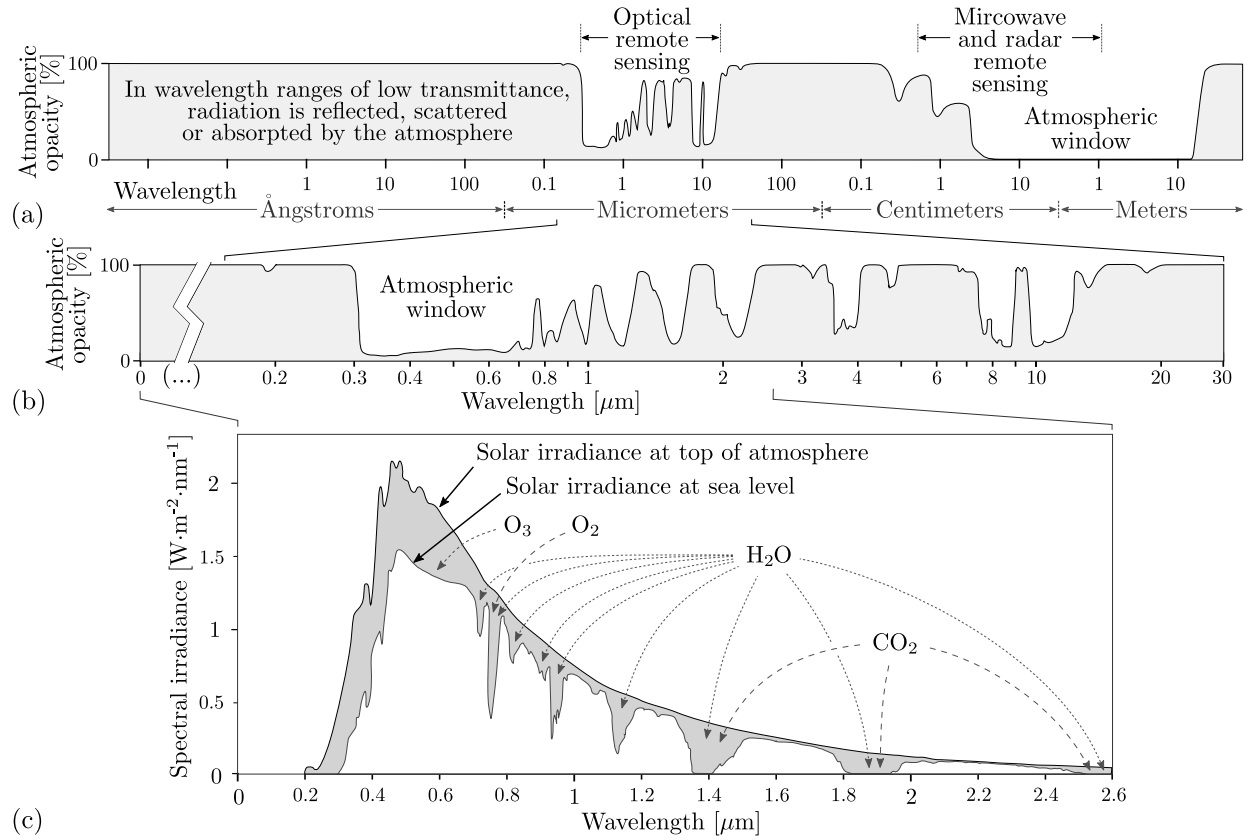


Fig. 3. Atmospheric opacity to electromagnetic energy in the wavelength regions (a) from less than 1 Ångstrom to 100 m and (b) from 0 to 30 μm. Mainly absorption, but also reflection and scattering processes lead to low transmittance of – i.e. high opacity to – radiation in most of the electromagnetic spectrum. The remaining spectral regions are characterized by "atmospheric windows" which remote sensing systems are restricted to. The heterogeneous absorption features in the optical wavelength range are mainly caused by various atmospheric gases including O_2, O_3, CO_2 and H_2O. (c) Compared to the measurable solar irradiance at the top of the of the atmosphere, the amount of solar irradiance reaching the Earth's ground at sea level is visibly reduced. *(Graphics and caption contain cumulative information originating from Slater [1980] and Jensen [2015, p. 189])*

◇ E_g is the total irradiance illuminating the observed ground surfaces.

The latter quantity can be further specified by the atmospheric transmittance, τ_θ, at incidence angle θ, the solar irradiance above the atmosphere, E_a, the sky irradiance E_d (constituted by the sky irradiances 1 & 2 shown in Fig. 1), and a factor F which accounts for the terrain topography, as follows:

$$E_g = \tau_\theta \cos(\theta) E_a + F E_d \tag{3}$$

At this point, the most important components of the at-sensor radiance, L, its interaction with the atmosphere and the Earth's surface, as well as its dependency on the material-dependent surface reflectance, ρ, are sufficiently comprehensively described and summarized by Eq. (2). But before moving on to the next topic, it is fundamental to notice that most of the terms introduced above strongly depend on two factors which have been merely foreshadowed up to this point, to facilitate readability. Those factors are the location of the observed area on the ground, given in 2-D coordinates (u, v), on the one hand, and the wavelength, λ, on the other hand. This omission is made up leeway now. Rewriting Eq. (2) with emphasis on these parameters lead to

$$L(\lambda, u, v) = \rho(\lambda, u, v) \frac{\tau_\phi(\lambda)}{\pi} E_g(\lambda, u, v) + L_p(\lambda) + L_e(\lambda, u, v). \tag{4}$$

With this dependency made explicit, Eq. (4) reveals the evident influence of the spectral wavelength on the signal and information acquired by a remote sensing system. As already broached at the beginning of Section 2.1, the sensitivity to specific wavelength ranges is also one of the main characteristics used to classify remote sensing systems.

Given the importance of the EM spectrum in this context, it is depicted in Fig. 2 with indication of the visible and infrared portions of the spectrum. The optical remote sensing interval ranges from around 0.4 to 15 μm and, therefore, covers the visible, as well as the near-, shortwave-, midwave- and longwave infrared spectral regions. This interval can be further divided into two parts, separated between the shortwave infrared (SWIR) and the midwave infrared (MWIR) at 3 μm, by the aspect of the remote sensor measuring *reflected* solar radiation (0.4 to 3 μm) or radiation *emitted* by the Earth's surface (3 to 15 μm). While the acquisition of data in the former (reflective) portion is primarily restricted to "daylight",

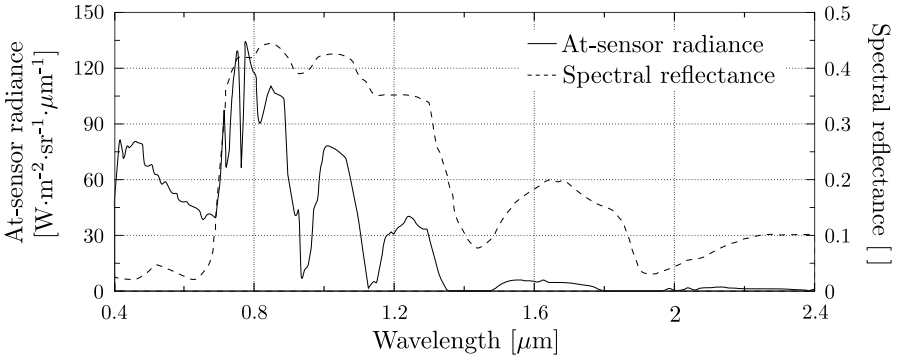


Fig. 4. Comparison of the at-sensor radiance spectral profile of grass, as measured by a satellite remote sensing system, and the corresponding reflectance spectral signature. The unitless reflectance values of a material are commonly specified in the unit interval, while at-sensor radiance values depend on the sensor design and its properties such as the radiometric resolution. Both the solar irradiance curve and the atmospheric absorption features are clearly visible in the at-sensor radiance profile, while the reflectance spectral signature solely depends on the wavelength-dependent reflectivity of the corresponding material, i.e. grass. *(After Schowengerdt [2006, p. 57])*

the latter portion corresponds to thermal energy which is emitted both during the day and at night. Since the overall surface temperature at night is generally lower than during the day, individual heat sources are more easily detected – i.e. discriminated from the background – by thermal scanners at night. It should be noted that, although the majority of optical remote sensing systems are *passive*, there are indeed *active* systems operating in the optical portion of the spectrum. *Light Detection And Ranging* (LiDAR) scanners, for instance, can be used to create 3-D point clouds by measuring distances. Airborne LiDAR is used e.g. to generate very high resolution digital elevation models (DEMs) or digital surface models (DSMs). There are many other applications for LiDAR scanners including atmospheric physics, agriculture, bathymetry and urban mapping to name just a few of many. However, since LiDAR systems are not subject of this thesis, they are no further extended on here.

The dependency of remote sensing systems on the operating wavelength range becomes ultimately evident when plotting the above-described overall influence of the atmosphere in terms of opacity over the wavelength. Such a graph is depicted in Fig. 3 (a). One can see that the atmosphere is fully opaque to radiation in most of the spectrum between 10^{-2} Å and 100 m. Those regions in which the atmosphere is transparent – the so called *atmospheric windows* – are what remote sensing systems are restricted to. Fig. 3 (b) shows a magnified view on the optical atmospheric window. The heterogeneous profile is shaped by atmospheric absorption features which are primarily caused by the presence of gases in the atmosphere, including O_2, O_3, CO_2, H_2O and N_2O. A detailed view on the solar irradiance measurable both above the atmosphere and at see level in the wavelength range from 0 to 2.6 μm is depicted in Fig. 3 (c). Furthermore, it indicates direct correspondences between the aforementioned gases and specific absorption features. This viewgraph explains why the *reflective* spectral region ends around 3 μm. Most of the solar irradiance reaching the Earth travels at wavelengths within this spectral portion. Under the assumption that the Sun is an ideal black body, this phenomena is described by Planck's law [Richards, 2013, pp. 32-33].

After having introduced the at-sensor radiance and surface reflectance, extended on the influence of the atmosphere, and explained the strong dependency of the remote sensing process on the wavelength, it is worth to illustrate how the at-sensor radiance and surface reflectance, corresponding to the same surface material, vary with the wavelength. This direct comparison is depicted in Fig. 4 exemplarily for Kentucky Bluegrass [Schowengerdt, 2006, p. 57]. Both the solar irradiance curve and the atmospheric absorption features, that are shown in Fig. 3 (c), are clearly visible in the at-sensor radiance profile, while the reflectance spectral signature solely depends on the wavelength-dependent reflectivity of the corresponding material.

Imaging Principles and Scanning Techniques Each pixel value in the acquired data can be thought of as the integral of radiation over a relatively small volume element in a 3-D continuous parameter space, $\mathbb{R}^3_+$, spanned by two spatial coordinates (u, v) and the spectral wavelength (λ). Fig. 5 illustrates such volume elements comparatively for the 10-m bands of the current Sentinel-2A space-based instrument and the visible to near infrared (VNIR) sensor of the upcoming EnMAP satellite mission. As it will be discussed later in this section, this parameter space is generally not as neatly divided, but "pixels" may overlap in both spectral and spatial directions. In most of the former and current optical satellite remote sensing systems, one of the two spatial dimensions in the image corresponds to the along-track direction, by means of the direction of motion of the sensor platform, while the other spatial dimension corresponds to the across-track direction which is approximately perpendicular to the along-track direction. This is a direct consequence of the interaction between the platform motion and the scanning procedure as will be explained in the following.

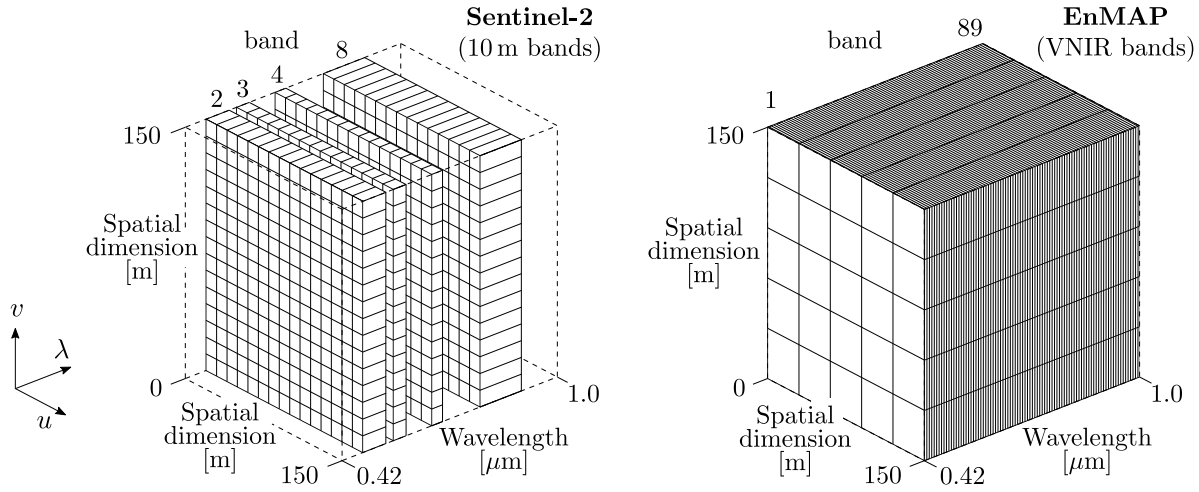


Fig. 5. Principal three-dimensional spatia-spectral representation of optical remote sensing data acquired by the 10-m bands of the multispectral Sentinel-2A imager (left) and the VNIR bands of hyperspectral EnMAP (right) mission. While the Sentinel-2A bands 2,3,4 and 8 have a finer spatial grid of 10 m, they are clearly outnumbered by the narrow 88 VNIR channels of EnMAP. *(Graphics inspired by Schowengerdt [2006, p. 19])*

Two common scanning techniques are depicted in Fig. 6 (a) and (b). Pushbroom scanners comprise a linear array of detector elements – the focal plane array (FPA) – aligned across-track, which scans the full swath simultaneously as the platform moves. This principle is shown in Fig. 6 (a). Whiskbroom scanners, on the other hand, scan in across-track direction, which is usually realized using an oscillating mirror, as shown in Fig. 6 (b). For both types of scanners, the full angular across-track coverage is referred to as the *field of view* (FOV). Hence, the corresponding swath width is also referred to as the *ground-projected field of view* (GFOV). The FPA can be 2-D as e.g. in the cases of staggered detector configurations or hyperspectral imagers. In the latter case, the additional dimension is used to have one FPA line for each spectral channel. Moreover, hyperspectral instruments are equipped with a dispersive prism or grating element, which disperses the incident radiation by wavelength, and is mounted between the imaging optics and the FPA. Along this second (spectral) dimension, the detectors are designed and calibrated to be sensitive only in the corresponding wavelength ranges, that is, only to the energy eventually reaching the individual detectors. Multispectral sensors without a dispersive optics use other technologies to discriminate different energy portions. Stripe filters e.g. are used in the multispectral instrument onboard the Sentinel-2 mission [Drusch et al., 2012].

2.1.2 Resolutions – PSF, SRF and SNR

Spatial Resolution and PSFs The spatial resolution of an optical imaging system is not uniquely defined. Broadly speaking, there are two conventional approaches to this definition: In electro-optical remote sensing imaging systems, the spatial resolution of the sensor (or its image) is often loosely quoted as either of the ground sampling distance (GSD) or the ground-projected instantaneous field of view (GIFOV). A simplified geometric description of those two terms is depicted in Fig. 7.

In photographic film cameras, on the other hand, the term “system’s spatial resolution” is commonly used to describe the minimum spacing between two just distinguishable lines on the film [Lillesand et al., 2014, pp. 136-140,223-224]. When projected onto the sensed surface, this spacing is referred to, synonymously, as the *ground resolution distance* or *ground resolved distance* (GRD). Fig. 8 illustrates the discrepancy of those two definitions by relating the GRD and the GSD. In order for two lines to be distinguishable in an image, there must

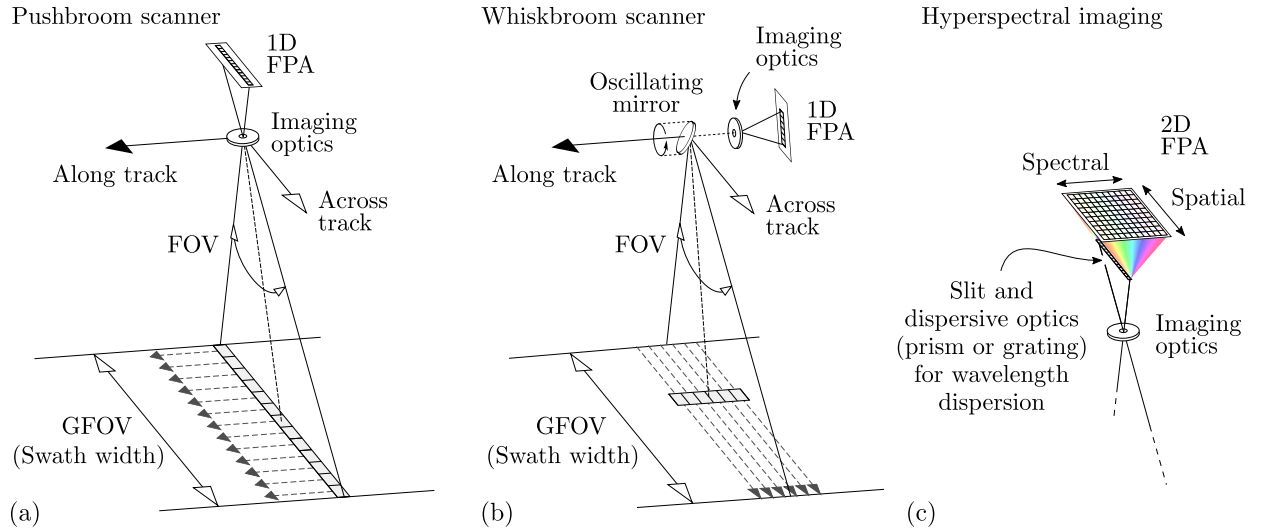


Fig. 6. Basic scanning and imaging principles, exemplified by the commonly used pushbroom (a) and whiskbroom (b) scanning techniques. Sub-figures (a) and (b) depict the geometric parameters and system components from the ground-projected field of view (GFOV) up to the focal plane array (FPA). (c) In the case of hyperspectral imagers, the incident radiation is spectrally dispersed via e.g. a prism or grating element before hitting the FPA. Compared to multispectral (panchromatic) imagers, for which multiple (one) one-dimensional FPAs principally suffice, hyperspectral instruments are commonly equipped with a two-dimensional FPA which detects incident energy in an array of detectors each of which is designed to have its peak spectral sensitivity in a specific wavelength range. *(Graphics and caption contain cumulative information originating from Schowengerdt [2006, pp. 20,28], Chuvieco and Huete [2009, pp. 74-77], Eismann [2012, pp. 243-246,314], Lillesand et al. [2014, pp. 220-225], and Richards [2013, pp. 6-8])*

be at least one pixel separating the lines. Hence, the GRD must be strictly greater than the GSD. According to Lillesand et al. [2014, p. 224] the optimal GRD achievable with a digital imaging system is approximately twice its GSD.

Having said this, there is yet another – more accurate – way of defining the system's spatial resolution by means of the widths of its *point spread functions* (PSFs). This definition will turn out to have certain advantages over the two previously introduced definitions, when it comes to multiresolution image fusion.

Before being converted to a digital number, the measured analog signal passes through multiple system components including the imaging optics, detectors and electronics. Each of these parts cause a spatial blurring of the incoming signal $s_b : \mathbb{R}^2 \to \mathbb{R}$ in band b [2], which images the continuous scene in the IFOV to an analog signal corresponding to one detector element in the focal plane at a specific time. This accumulated blurring effect is characterised by the system's spatial responsivity, which can be quantified by its net (total) PSF, PSF. The resulting electronic signal, e_b, is given by the convolution

$$e_b(u,v) = \int_{\alpha_{\min}}^{\alpha_{\max}} \int_{\beta_{\min}}^{\beta_{\max}} s_b(\alpha,\beta) PSF(u-\alpha, v-\beta)\, d\beta d\alpha\,, \tag{5}$$

where the limits in the integrals are defined by the PSF's spatial extend around the coordinates (u,v), which, in turn, represent the center of one detector element or, if projected onto the ground, the center of the ground resolution cell. For the types of sensors discussed in this work, the PSFs can be approximated by separable ones [Schowengerdt, 2006, p. 86]. Moreover, it can be shown that PSF is expressible by the convolution of individual system

[2] The concept of spectral bands will be introduced later in this subsection.

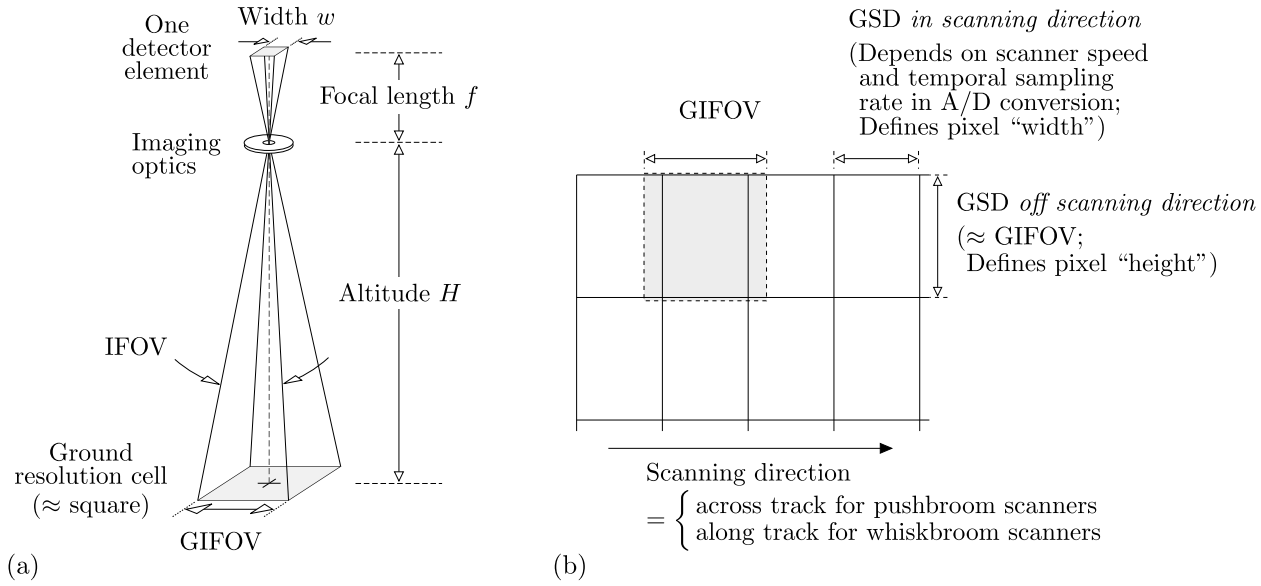


Fig. 7. (a) Simplified geometric description of the relationship between a single detector element, its instantaneous field of view (IFOV) and its altitude-dependent ground resolution cell. In reality, $f \ll H$, $w \ll$ GIFOV, and the effective ground resolution cell is rather elliptical than rectangular and somewhat distorted as a result of off-nadir viewing angles and motions during the signal integration time in the analog-to-digital (A/D) conversion. (b) The ground sampling distances (GSDs) along and across the track are generally not identical. Along the scanning direction, it primarily depends on the scanning speed (platform motion for pushbroom and mirror oscillation speed for whiskbroom systems) and on the temporal sampling interval used in the A/D conversion. In the case of pushbroom scanners, perpendicular to the scanning direction, the GSD depends on the size and inter-detector spacing along the FPA, which do not depend on motion. *(Graphics and caption contain cumulative information originating from Schowengerdt [2006, pp. 21-22] and Lillesand et al. [2014, pp. 223-225])*

component PSFs [Schowengerdt, 2006, pp. 85-91]:

$$PSF(u,v) = PSF_{\text{opt}} * PSF_{\text{det}} * PSF_{\text{mot}} * PSF_{\text{el}}(u,v) \tag{6}$$

Those PSF components are briefly described in the following:

Optical PSF_{opt} The PSF_{opt} component corresponds to the spatial energy distribution in the image of a point source. Its spreading depends on factors including aberrations, optical diffraction and the quality of the system elements, especially, the FPA [Eismann, 2012, pp. 250-256,266-285]. Despite of its complexity, this part is typically modeled as simple 2-D Gaussian function

$$PSF_{\text{opt}}(u,v) = \frac{1}{2\pi ab} \exp\left(-u^2/a^2\right) \exp\left(-v^2/b^2\right), \tag{7}$$

where a and b determine the PSF's width in along- and across track direction, respectively.

Detector PSF_{det} This component models the spatial limits of the square detector element:

$$PSF_{\text{det}}(u,v) = \begin{cases} 1 & \text{if } \max\{|u|,|v|\} \le w/2 \\ 0 & \text{else} \end{cases} \tag{8}$$

Image Motion PSF_{mot} During the signal integration time, Δt, the scanner movement causes a 1-D smearing along the scanning direction, i.e. across track for whiskbroom and along

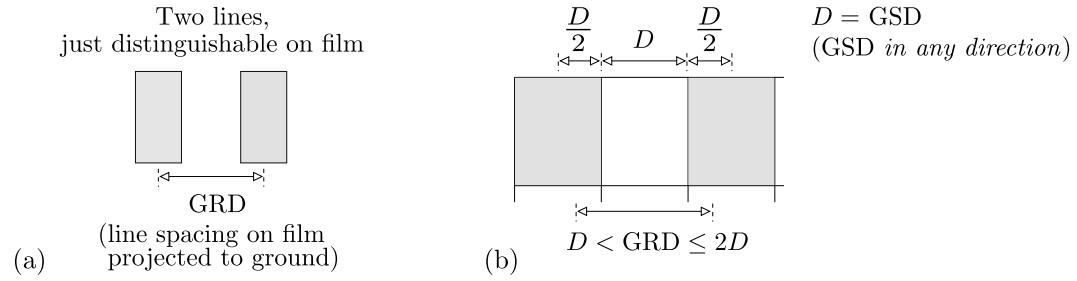


Fig. 8. (a) The spacing between two lines on the sensed surface, that are just distinguishable on a camera's film, is called the ground resolution distance (GRD). (b) Relationship between the GRD and the GSD. Both the GRD and the GSD are commonly referred to as *spatial resolution* in digital image processing. However, as illustrated in this sub-figure, those two quantities are not identical and should, therefore, be used consciously. *(After Lillesand et al. [2014, p. 224])*

track for pushbroom scanners:

$$PSF_{\text{mot,whiskbroom}}(u,v) = \begin{cases} 1 & \text{if } |u| \leq s/2, \text{ where } s = \Delta t \cdot \text{scan velocity} \\ 0 & \text{else} \end{cases} \tag{9}$$

$$PSF_{\text{mot,pushbroom}}(u,v) = \begin{cases} 1 & \text{if } |v| \leq s/2, \text{ where } s = \Delta t \cdot \text{platform velocity} \\ 0 & \text{else} \end{cases} \tag{10}$$

Electronics PSF_{el} The electronics components may cause additional blurring and are sometimes used to intentionally filter the signal over the integration time to reduce noise. As the imaged scene "moves" during the integration period, this temporal filtering can be converted to spatial blurring mainly along the scanning direction. As this component is often neglected in system models including the ones used in this thesis, the PSF_{el} component is not further discussed here.

With the definition of the net PSF, we are now able to bring the PSF into the context of spatial resolution as it is described above via the GIFOV: An electro-optical scanning system is often designed and calibrated so that its GIFOV is approximately equal to the ground-projected PSF's *full width at half maximum* (FWHM) [Alparone et al., 2015, p. 18]. This relationship, as well as the visual explanation of FWHM, are depicted in Fig. 9. Note, that the GIFOV is not necessarily the same in along- and across-track direction, and that the *PSF* is assumed to be separable, but not necessarily direction invariant. Nonetheless, the above-described relationship is commonly assumed to hold in both spatial dimensions.

At this point, is it expedient to introduce how Gaussian distributions can be expressed in terms of FWHM. This definition is commonly used in photo-electron spectroscopy and other fields in which Gaussian curves describe physical energy distributions [Gao et al., 2015] [Meikle, 2008, p. 47]:

$$I(E) = h \exp\left(-4\ln(2)\frac{(E-c)^2}{\text{FWHM}^2}\right) \tag{11}$$

In Eq. (11), $I(E)$ describes the accumulated measurable intensity, E represents the energy, h is the Gaussian curve's peak height and c the center (peak) position. Moreover, the vertical scaling factor, h, can be indirectly expressed via the area, A, under the curve:

$$h = \sqrt{\frac{\ln(2)}{\pi}}\left(\frac{2A}{\text{FWHM}}\right) \tag{12}$$

Both alternatives are used in this thesis, as they allow for a simple normalization of the

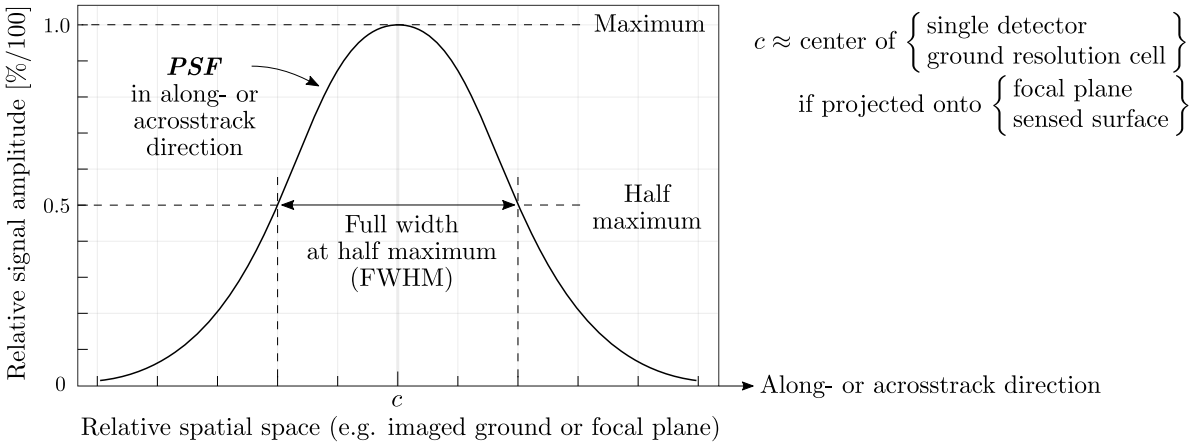


Fig. 9. Illustration of a modeled system's optical PSF in one direction (along- or acrosstrack) and its full width at half maximum (FWHM). Electro-optical scanning systems are often be designed and calibrated such that its ground-projected PSF's FWHM is approximately equal to the GIFOV [Alparone et al., 2015, p. 18].

curve by height and area.

To close the circle: In preliminary and feasibility studies, it is common practice to approximate the PSF around the ground resolution cell's center point (u_c, v_c) by using a Gaussian function in conjunction with the its FWHM which, in turn, is sometimes substituted by the GIFOV or the GSD depending on how much system information is available. In this case, Eq. (6) simplifies to

$$PSF(u,v) = \sqrt{\frac{\ln(2)}{\pi}}\left(\frac{2A}{\mathrm{FWHM}}\right)\exp\left(-4\ln(2)\frac{(u-c_u)^2}{\mathrm{FWHM}^2}\right)\exp\left(-4\ln(2)\frac{(v-c_v)^2}{\mathrm{FWHM}^2}\right) \tag{13}$$

This model was used in the experiments presented in Sections 4 and 5 wherever there was no further knowledge available about the systems' PSFs. The theoretical part, however, is developed in terms of non-modeled PSFs in order to preserve generality.

Before stepping on from spatial to spectral responsivity, there are some remarks on spatial resolution and PSFs worth mentioning for the sake of completeness:

- ◇ Note, that spectral dependences of the parameters occurring in the definitions of the system's spatial response and resolution have been widely ignored in this section for simplicity. In reality, PSFs not only vary between the detector elements in the same band, but also between spectral bands.
- ◇ The system's spatial responsivity can be analogously described in terms of *modulation transfer functions* (MTFs) instead of PSFs, where an MTF is the magnitude of the Fourier transform of the PSF [Solomon and Breckon, 2011, pp. 131-134], [Schowengerdt, 2006, pp. 257-260], [Eismann, 2012, p. 256].
- ◇ For further reading on spatial resolution and responsivity, the interested reader is encouraged to refer to Alparone et al. [2015, pp. 15-21], Gonzalez and Woods [2006, p. 81], Schowengerdt [2006, pp. 77-82,85-103], Jensen [2015, pp. 14-18], Lillesand et al. [2014, pp. 136-142], Eismann [2012, pp. 247-266,300-307] and Richards [2013, pp. 142-144].

Spectral Resolution and SRFs The spectral characteristics of an electro-optical remote sensing system is determined by a number of factors including the sensor's sensitivity to specific wavelengths and the number and specification of spectral *channels* or *bands*. As mentioned above, multiband imaging can be realized by using e.g. a dispersive optics, such as a prism or grating element, in combination with a 2-D FPA whose detectors are calibrated

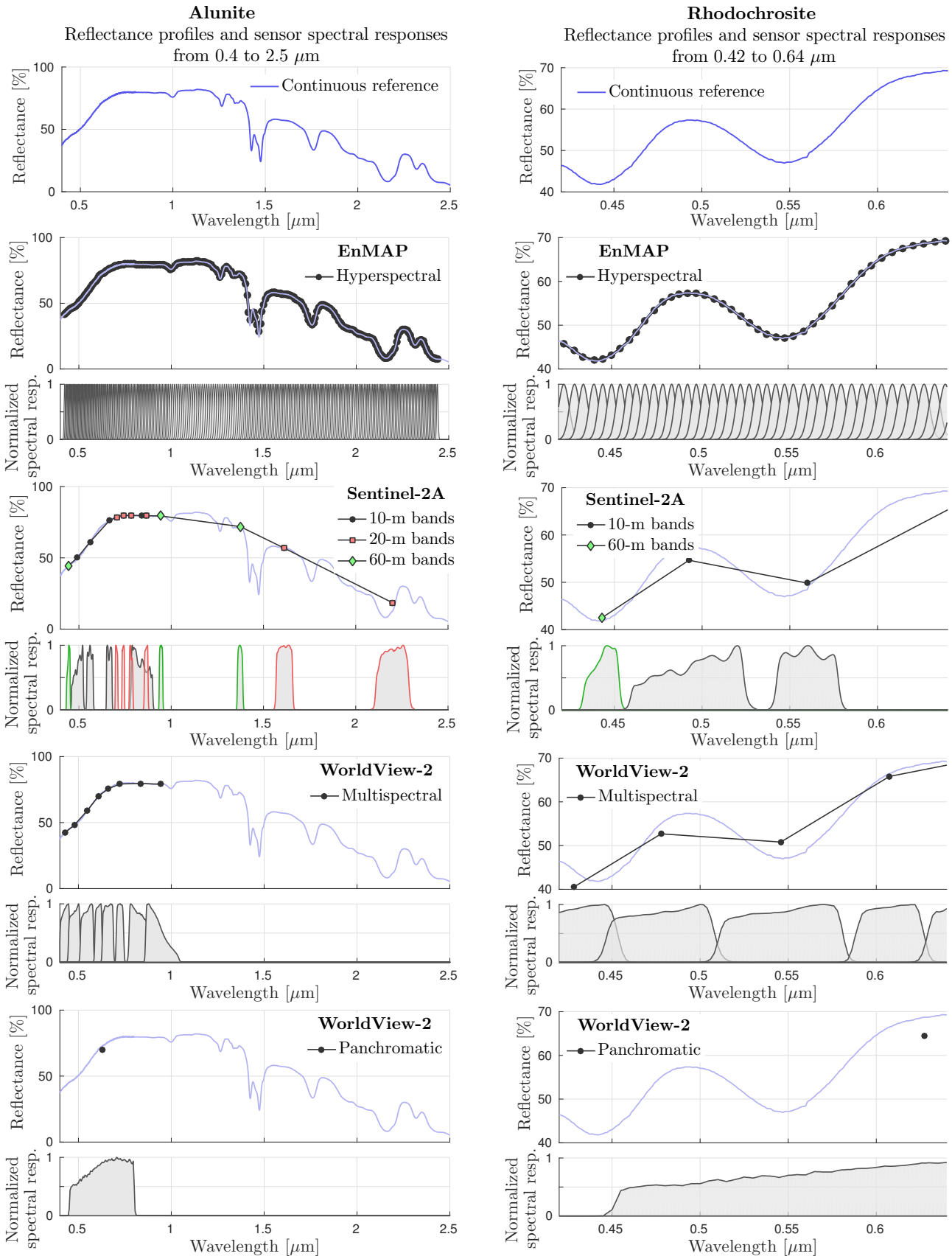


Fig. 10. Spectral reflectance signatures of the two minerals Alunite (left column) and Rhodochrosite (right column), and corresponding simulated spectral reflectance profiles acquirable by selected current and future optical satellite remote sensing systems of different spectral characteristics. For every sensor, the normalized SRFs are plotted. The SRFs, and even more so the differences between the discrete measurable reflectance profiles (black dotted curves) and the smooth spectral signatures (blue curves), indicate the sensors' spectral mapping capacity, i.e. resolution. Noticeable, while EnMAP is capable of accurately measuring the absorption features of both minerals, the broadband Pan SRF of WorldView-2 merely filters signals to one number and is, hence, not particularly useful for material identification purposes.

to be spectrally sensitive to specific wavelength ranges (see Fig. 6). If a sensor is designed to have one spectral channel only, it is synonymously called a *panchromatic* or *monochromatic* sensor. Those instruments are usually sensitive to a broad wavelength interval within the visible or visible to near infrared (VNIR) portion of the spectrum. Multiband sensors can be further categorized into multi-, super-, hyper- or ultraspectral sensors, depending on the number of bands and the band's widths. To date, there is no standard defining clear boundaries between these category. Alparone et al. [2015, pp. 22-23] speak of

- ◇ multispectral, if the number of bands lies between 3 and 10,
- ◇ superspectral, if the number of bands exceeds 10,
- ◇ hyperspectral, if the bandwidths are narrower than or equal to 10 nm, and
- ◇ ultraspectral, if the bandwidths are narrower than or equal to 1 nm.

In this work, multi- and superspectral sensors are equally referred to as multispectral sensors, and hyper- and ultraspectral sensors are equally referred to as hyperspectral sensors. Independent of the type of the sensor, each band is characterised by its spectral responsivity, which is quantifiable by a *spectral response function* (SRF) over the wavelength. The SRF describes the spectral distribution of energy detectable by one detector line of the FPA, i.e. by all detectors that correspond to the same spectral band [3] [Eismann, 2012, p. 341]. From a signal processing point of view, most of the sensor spectral characteristics can be derived from the SRFs.

Similarly to the definition of the sensor's spatial resolution, also the spectral resolution is not uniquely quantifiable, but rather context-dependent. One general yet accurate formulation defines spectral resolution as the sensor's capability to resolve details in the spectral signal [Schowengerdt, 2006, p. 82]. Another common definition states that the spectral resolution refers to the number of spectral bands *as well as* their bandwidths [Chuvieco and Huete, 2009, pp. 65-66], [Jensen, 2015, p. 12]. Under the assumption of a reasonable sensor design, the latter implies the former definition to some extent, because the more and the narrower the bands, the better the discrimination capacity of the sensor.

The inevitable question arising from either definition is: What exactly is it, the sensor is – or, if not built yet, should be – capable of discriminating? A general answer to this question would be *materials*, or "objects on the ground". Moreover, discrimination is certainly the less challenging, the more accurately materials are individually *identifiable* by the sensing system. The discrimination, and even more so the identification, of materials are best done in the reflectance domain. This is because the acquired – or, more precisely, reconstructed – spectral reflectance profile of one spatial pixel can be directly compared to *spectral reflectance signatures*, which are smooth – or nearly smooth – functions of wavelength representing materials' relative surface reflectivity independently of the solar irradiation.

An important factor in both material discrimination and identification is the knowledge of the significant *features* in the material's spectral signature. Those features are particularly distinct in narrow wavelength ranges in which the material is strongly absorptive. In this case, they are referred to as *absorption features*.

Fig. 10 shows the spectral reflectance signatures of two minerals and corresponding spectral reflectance profiles acquirable by various current and future optical satellite remote sensing systems of different spectral characteristics. It also shows the SRFs of the corresponding sensors. Together with the SRFs, the signals acquirable by the sensors give rise about the

[3] More precisely, every single detector within the FPA has an individual SRF. The SRF associated with one band is the average of all SRFs that correspond to detectors belonging to the same band. Minimizing the standard deviation of those individual SRFs to the average band SRF is subject of the detector design and sensor calibration phase.

sensors' spectral capabilities. The more precisely the sensor is capable of mapping spectral profiles, the better is it's material identification – and, thus, discrimination – capacity.

The absorption features of Alunite lie around 1450, 1750 and 2200 nm. From the four sensors under comparison, only the hyperspectral EnMAP sensor is capable of accurately measuring these features and hence detecting Alunite in a scene with a high certainty. A similar conclusion can be drawn if comparing the two absorption features of Rhodochrosite (around 440 and 550 nm) in the reference spectral signature with the corresponding simulated sensor signals. The extreme counterpart to hyperspectral sensor is a panchromatic sensor like the one of WorldView-2. The plots on the lower right of Fig. 10 show how the broad SRF of the panchromatic sensor "smoothes out" all spectral details to produce a single number. Such single-band pixel values are doubtlessly insufficient to identify materials.

After EnMAP, Sentinel-2A has the second finest spectral resolution in this comparison, and is what Alparone et al. [2015] refer to as a superspectral sensor. Especially Sentinel-2A's four narrow adjacent 20-m GSD bands around the red-edge spectral region indicate high discrimination capabilities of materials whose spectral signatures are particularly distinct in this wavelength range. These materials include almost all kind of "green" vegetation. Furthermore, in comparison to WorldView-2, Sentinel-2A has additional bands beyond the near infrared.

The WorldView-2 satellite carries both a multispectral and the above-mentioned panchromatic sensor. The SRFs of the multispectral sensor range over the visible to near infrared spectral portion, which is more typical for classical optical multispectral scanning systems.

Crucially, higher spectral resolution is inevitably traded by degraded spatial resolution. Among the four sensors compared in Fig. 10, EnMAP has the coarsest spatial resolution of 30-m GSD, followed by Sentinel-2A with 10-m GSD at best. Next, the multispectral bands of WorldView-2 provide 2-m GSD imagery, and lastly, WorldView-2's panchromatic sensor has the highest nadir spatial resolution of 0.5-m GSD. This evident trade-off between spectral and spatial resolution is of substantial importance and one of the main driving factors for the work presented in this thesis.

More information about specific missions, sensors and their characteristics will be provided and discussed in Section 2.1.7.

Relationship Between Digital Number, At-sensor Radiance, Reflectance, PSF and SRF With R_b denoting the SRF corresponding to band b, we are now able to bridge the gab between Eq. (4) and Eq. (5); That is, to relate the spectrally and spatially filtered continuous electronic signal, e_b, corresponding to one detector element in the at-image plane (before the A/D conversion to a digital number), to the at-sensor radiance, L. Therefore, it is auxiliary to introduce the spectral irradiance, E_i, on a detector by the simplified *camera equation* [Slater, 1980; Schowengerdt, 2006, p. 104]:

$$E_i(\lambda, u, v) = \frac{\pi \tau_o(\lambda)}{4N^2} L(\lambda, u, v), \tag{14}$$

In Eq. (14), τ_o denotes the optical system transmittance and N the optics' f-number. The as yet unspecified signal s_b appearing in Eq. (5) can now be explicitly expressed as follows:

$$s_b(\lambda, u, v) = \int_{\lambda_{\min}}^{\lambda_{\max}} R_b(\lambda) E_i(\lambda, u, v)\, d\lambda \tag{15}$$

Substituting Eq. (4), (14) and (15) into Eq. (5) leads to

$$e_b(u,v) = \int_{\lambda_{\min}}^{\lambda_{\max}} \int_{\alpha_{\min}}^{\alpha_{\max}} \int_{\beta_{\min}}^{\beta_{\max}} \overbrace{R_b(\lambda)\,\frac{\pi\tau_o(\lambda)}{4N^2}}^{\textbf{SRF}} \overbrace{\left(\overbrace{\rho(\lambda,\alpha,\beta)}^{\textbf{Reflectance}}\,\frac{\tau_\phi(\lambda)}{\pi} E_g(\lambda,\alpha,\beta) + L_p(\lambda) + L_e(\lambda,\alpha,\beta) \right)}^{\textbf{At-sensor radiance } (=L(\lambda,\alpha,\beta))} \overbrace{PSF(u-\alpha, v-\beta)}^{\textbf{PSF}}\, d\beta\, d\alpha\, d\lambda\,. \tag{16}$$

The only term yet to be derived is the digital number, *DN*. Therefore, the electronic signal, e_b, is amplified and shifted by an electronics gain, $gain_b$, and offset, $offset_b$ in band b, respectively, before being sampled and quantized into *DN*s in the A/D step. If p is the pixel in the image associated with the detector element, which has its center point at (u,v), and $\lfloor \cdot \rceil : \mathbb{R} \longrightarrow \mathbb{Z}$ denotes the simple quantizer rounding a real number to the closest integer number, then the digital number of pixel p in band b can be written as

$$DN_{b,p} = \lfloor\, gain_b \cdot e_b(u,v) + offset_b \,\rceil. \tag{17}$$

The amplification of e_b is important to make the maximum use of the predefined full digital number range in the A/D conversion. The gain and offset are set during the sensor calibration stage by estimating the expected maximum at-sensor radiance range and, correspondingly, the lower and upper bound for e_b. Occasional occurrences of saturation under exceptional illumination conditions during the scanning over highly reflective materials are tolerated to improve the overall gain, and hence the radiometric resolution [Schowengerdt, 2006, pp. 106-107] as will be discussed in the following paragraph.

Radiometric Resolution and SNR The radiometric resolution describes the precision with which intensity (brightness) values can be digitally mapped. The larger the effectively used discrete number range for *DN*s and the higher the *signal-to-noise ratio* (SNR) (see description below), the better the digital representation precision of the analog signal. As was discussed in the previous paragraph, the parameters $gain_b$ and $offset_b$ are used to amplify the electronic signal e_b in band b to the full potential number range of *DN*s. This allows setting the lower bound of this *dynamic range* to 0. Hence, the upper bound plus one defines the effectively used number range, i.e. the radiometric resolution. In order to avoid memory wastage, the only reasonable choice for this range is $[0, 2^Q - 1] \subset \mathbb{Z}$, where Q is some – usually even – positive integer which is limited by the system's SNR. This fact makes the radiometric resolution expressible by the number of *bits per pixel*. Simple digital RGB cameras, for instance, have an 8-bit radiometric resolution, which means that all pixels take integer values (*DN*s) between 0 and 255. This level of precision is sufficient for imagery which is meant to be observed by the human eye only. If more sophisticated information needs to be extracted or derived from the data, a finer radiometric resolution might be indispensable, which is why optical satellite remote sensing systems usually have a dynamic range of more than 10 bit.

The SNR is a unitless measure to quantify the contribution of the pure *signal*, which is the unknown noiseless part of the measurement carrying the information of interest, relative to the undesired *noise*, which is, unavoidably, cumulatively introduced by all components

of the sensing system as well as all media the energy travels through. There are multiple inter-convertible formulations of the SNR based on either the *amplitude* or the *standard deviation* of the signal and, correspondingly, the noise. If σ_x denotes the standard deviation of x and C_x the *contrast* of x, which is given by either the ratio of or the difference between the maximum and the minimum value of x, then the definitions of the SNR commonly found in the literature include the following [Schowengerdt, 2006, pp. 146-149], [Raol, 2015, p. 254], [van der Meer and de Jong, 2013, p. 35], [Bieniarz, 2015, pp. 21-22]:

$$SNR_{\text{power}} = \left(SNR_{\text{amplitude}}\right)^2 = \left(\frac{C_\text{s}}{C_\text{n}}\right)^2 \tag{18}$$

$$SNR_{\text{var}} = \left(SNR_{\text{std}}\right)^2 \qquad = \left(\frac{\sigma_\text{s}}{\sigma_\text{n}}\right)^2 \tag{19}$$

Since the maximum and minimum values of the signal, and therefore C_s, are highly vulnerable to outliers caused by e.g. dead pixels, the definitions given in Eq. (19) based on the standard deviation are preferable from this aspect and, therefore, used in this work. Moreover, it is beneficial to transform any of the above-defined SNR terms to decibels (dB). The power and variance SNR terms, for instance, are transformed to dB as follows:

$$SNR_{\text{dB}} = 10\log_{10}(SNR)\,[\text{dB}] \tag{20}$$

Importantly, the as yet unspecified *signal* (and, correspondingly, the noise) can be given in any of the above-introduced data formats, i.e. reflectance, at-sensor radiance, or digital number.

Considering the number and complexity of all factors potentially contributing to the measurement noise, realistic noise models can be arbitrarily sophisticated. One comparatively comprehensive model is discussed in [Eismann, 2012, pp. 281-294]. In summary, the aggregate system noise comprises *signal shot-*, *dark-current-*, *Johnson-*, *readout-*, *spatial FPA-non-uniformity-*, and *A/D-noise* and is given by

$$\sigma_n = \sqrt{\sigma_{\text{n,shot}} + \sigma_{\text{n,dark}} + \sigma_{\text{n,Johnson}} + \sigma_{\text{n,read}} + \sigma_{\text{n,spatial}} + \sigma_{\text{n,A/D}}}\,. \tag{21}$$

In practice, the noise level is commonly estimated exclusively from the data. One simple method, for instance, considers one image patch within a bright homogeneous area and estimates the SNR as the mean value over the standard deviation of all pixels in the patch [Bieniarz, 2015, p. 22]. Other practical yet more substantial methods for noise estimation consider linear transformations, such as principal component analysis (PCA), knowing that the SNR is much higher in the leading than in the tailing principle components [Pyatykh et al., 2013; Liu et al., 2013].

Temporal Resolution Apart from the spatial, spectral and radiometric resolution, and the SNR, another system parameter is the *temporal resolution*, which is defined as the time interval between two scans of the same area on the ground. It is primarily determined by the satellite orbit, the swath width and the system's sensor pointing capabilities. Noticeably, the temporal resolution differs from the *orbital revisit frequency* in that the latter quantity defines the repetition time it takes for a satellite to observe the same spot *from the same orbit*. While the orbital revisit frequency does not take into account possible swath overlapping and pointing capabilities, the temporal resolution does and is, therefore, potentially shorter.

The temporal resolution strongly varies between remote sensing satellite missions according

to the missions' objectives. The design of meteorological satellites, for instance, is primarily driven by a high demand for highly frequent information updates as the observed phenomena are highly dynamic. Moreover, for the purpose of weather monitoring, very high spatial resolution would not only be little helpful, but make data handling, i.e. storage, transfer and processing, much more challenging as the data volume substantially increases. Other applications benefiting from high temporal resolution include change detection and the monitoring of hazardous events or specific regions or objects on the ground such as agricultural crops.

Mineralogical missions, in contrast, are driven by higher demands for high spatial resolution and, even more so, high spectral resolution. The temporal resolution is of minor importance.

2.1.3 Geometric Correction, Calibration and Noise Reduction

As we have seen from the introduction up to this point, there are more than one representation of optical Earth observation data. In order to be able to extract as much information from the data as possible, the raw data is processed according to a processing chain equal or similar to the one shown in Fig. 11. In addition to the reconstruction of at-sensor radiance and reflectance data, one important element is the geometric correction. Moreover, noise reduction can help to further enhance the quality of the data for maximum information accessibility. The ordering of the individual components and subcomponents in the processing chain is partially interchangeable. In particular, atmospheric correction is sometimes considered separately from radiometric correction. Depending on the application or general purpose of the data, not all correction and calibration steps may be required.

Organisations such as the DLR German Aerospace Center and the National Aeronautics and Space Administration (NASA) have developed standards for this processing chain which define the processing ordering and associate product levels with intermediate data products [Krauß et al., 2013; Müller et al., 2010; NASA, 1992].

In the following, the main steps in geometric correction, radiometric calibration and noise reduction are briefly described.

Geometric Correction In Earth observing remote sensing it is desirable to know what location on the ground each pixel corresponds to with respect a to ground-based coordinate system. There are a number of factors causing geometric distortions in the original data that need to be compensated for in order to improve interpretability and usefulness of the data. For satellite-based systems, those factors include orbital characteristics, platform attitude, rotation, shape and topography of the Earth, and sensor properties.

The correction process is conducted according to a suitable projection model and involves coordinate transformations as well as data resampling (interpolation) to a reference grid. Possible means of geometric correction include the following:

Rectification/Georeferencing: The planimetric alignment of the image to a map.
Geocoding: A special case of rectification with an additional resampling of all pixels to the same GSD.
Orthorectification: Consideration of the surface topography. This step requires auxiliary information such as a *digital elevation model* (DEM).
Registration: The alignment of the image to another image.

Furthermore, auxiliary information such as *ground control points* can be used to further improve the geometric accuracy. For more in-depth information on the various elements of

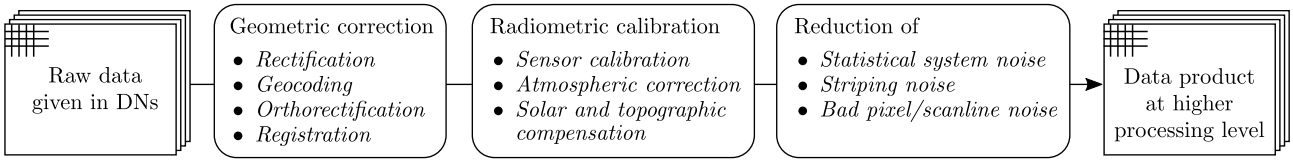


Fig. 11. Flowchart of a basic raw data processing chain. The ordering of individual components and subcomponents may vary between organisations and processing standards. In particular, atmospheric correction is sometimes considered separately from radiometric correction. Moreover, depending on the application or general purpose of the data, not all correction and calibration steps may be required.

geometric distortions and correction the interested reader is encouraged to refer to Jensen [2015, pp. 235-271], Schowengerdt [2006, pp. 110,286-308] and Chuvieco and Huete [2009, pp. 219-234].

Radiometric Calibration After having introduced the basic concept of how the digital raw data (DNs) of an optical remote sensing system is produced, this paragraph briefly summarizes the processing steps – by means of calibration – required to reconstruct the at-sensor radiance, surface radiance and surface reflectance. The representation of the data in physically meaningful units is not only beneficial for the analysis and interpretation of individual images acquired by one sensor, but can be essential if one wants to compare or even fuse the data of different sensors. Especially the individual gains and offsets of each sensor for the DN generation makes a comparison of DN data from different sensors basically meaningless. The overall correction for the influence of the sensor system, atmosphere, solar irradiance and terrain topography on the signal at different stages, is referred to as *radiometric calibration* [Schowengerdt, 2006, pp. 332-350]. Fig. 12 shows a flowchart of this calibration chain.

Sensor Calibration: The sensor calibration converts the raw DNs to at-sensor radiance values. This step principally requires solving Eq. (16) and (17) for $L_{b,p}$, which is the integral of the continuous at-sensor radiance, $L(\lambda, u, v)$ over the support of the SRF and the support of the PSF corresponding to pixel p in band b. In practice, this solution is usually approximated based on various simplifications. The corresponding coefficients are calculated before the satellite's operation phase and stored in *calibration tables*.

Atmospheric Correction: As described at the beginning of Section 2 and depicted in Fig. 1, the at-sensor radiance is strongly influenced by the atmosphere. This implies dynamic phenomena that need to be corrected for in order to calculate the surface radiance, and later the surface reflectance. Atmospheric correction models, methods and processors are still subject to research and development. Richter et al. [2011], for instance, recently proposed a method which – in addition to an established atmospheric modelling part – implicitly correct for the *smile* effect [4]. This method will be used in the processing and archiving facilities (PAFs) of the planned German hyperspectral EnMAP mission. Generally, atmospheric correction accuracy of a specific scene, observed at a specific time, can be significantly improved by taking ground measurements of the primarily influencing gases (see Fig. 3) in the atmosphere at the place and time of acquisition.

Solar and Topographic Compensation: In order to recover the surface reflectance value, $\rho_{b,p}$, corresponding to band b and pixel p, the atmospherically corrected data requires further correction for signal influences introduced by the solar irradiance and the terrain topography. As mentioned on page 7, the model derived in Section 2.1.1 is based on the assumption of Lambertian surfaces, which is a simplification of the reality allowing for a more intuitive introduction of the subject. In practice, surfaces reflection is partly directional and, hence, affected by terrain declinations. This influence of the topography

[4] The *smile* effect will be briefly introduced in Section 2.1.5.

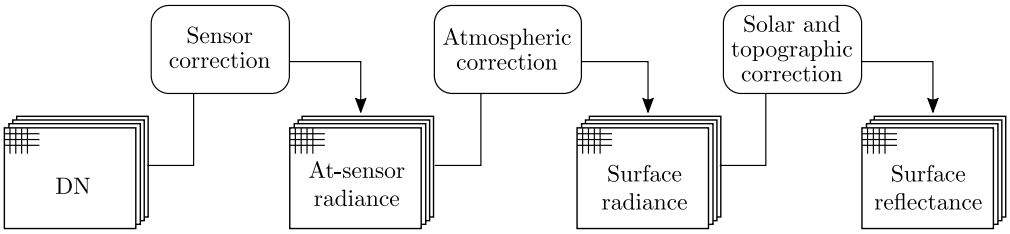


Fig. 12. Flowchart of a typical calibration chain used in optical remote sensing. Beginning with unitless raw digital numbers (DNs), sequential correction of the influence of the sensor system, atmosphere, solar irradiance and terrain topography on the signal at different stages, allows for the reconstruction of physically meaningful quantities. *(After Schowengerdt [2006, pp. 335])*

on the reflection behaviour can be compensated for using a DEM. Moreover, for the correction of the influence of the solar irradiance, the incident angles from Sun to surface and from surface to sensor are accounted for under consideration of the terrain inclination. Similarly to the atmospheric correction, also the solar correction can be improved if ground measurements of the solar irradiance from the observed scene are available.

Noise Characteristics and Reduction As mentioned in the introduction of the SNR further above in this section, there are many types of noise that can corrupt the signal. Noise sources can be any system component and any medium the signal travels through. Particularly prominent noise constituents include the signal *shot-*, *dark-* and *Johnson noise*, which accumulate to what is known as *detector noise* (compare Eq. (21)). While – due to the Poisson statistical nature of the quantum detection process – shot- and dark noise underlie Poisson distributions, the Johnson part is associated with thermal carrier motion which can be modeled as a signal-independent, zero mean, additive process [Eismann, 2012, pp. 281-284]. Not only because, at higher signal amplitudes, Poisson distributions can be approximated by a normal distribution with signal dependent-variance, but also because all other noise contributions in Eq. (21) are typically modeled as normal distributions, the total system noise is commonly described by a normal distribution as well [Eismann, 2012, p. 292] [Simões et al., 2015] [Loncan et al., 2015].

Due to both the complexity of the system noise constituents, and the inaccessibility of some instrument components after launch, the total system noise is often estimated statistically rather than forward-modeled. For panchromatic and multispectral imagery, one approach to estimating the system noise is to derive the noise statistics from the difference between a uniform area in the scene and a spatially slightly shifted version of it, while assuming that, in contrast to the pure signal component, the noise part is spectrally and spatially uncorrelated [Eismann, 2012, p. 519]. For hyperspectral imagery, it is common to specifically exploit the inherent mutual correlation between the spectral bands [Yokoya, Grohnfeldt, and Chanussot, 2017]. One simple yet effective method approximates each band by a linear combination of all *other* bands in the least squares sense, and considers the residual as the additive noise component in that band [Bioucas-Dias and Nascimento, 2008]. For noise reduction purposes, those residuals are then simply discarded.

Other types of noise include *striping noise* and *bad scanline noise*. Striping noise is primarily caused by the individual calibration of each detector element in a laboratory under non-operational conditions prior to the operational phase and can be corrected for independently of other noise components. One of many methods addressing this problem identifies the stripes based on edge detection and line-tracing algorithms and reduces the striping effect via cubic spine interpolation [Tsai and Chen, 2008]. Bad scanline noise usually results from a defective detector which creates a "bad" – or "missing" – line, i.e. salt and pepper values, in scanning direction as the sensor scans the scene. Remedy can be found, for instance,

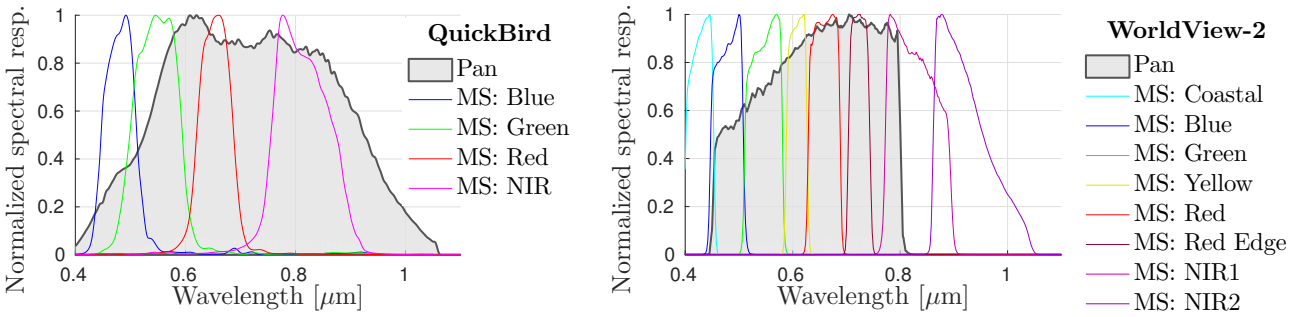


Fig. 13. Normalized spectral response functions of the QuickBird multispectral and panchromatic sensors (left) and the WorldView-2 multispectral and panchromatic sensors (right).

by interpolation [Chuvieco and Huete, 2009, pp. 194-195].

For a more comprehensive yet illustrative discussion on noise models and reduction the interested reader is encouraged to refer to Schowengerdt [2006, pp. 140-152, 315-328].

2.1.4 Specifics in Panchromatic and Multispectral Remote Sensing

Multispectral imagers are characterized by multiple – typically 3 to 10 – spectral channels each of which is spectrally responsive in a wavelength range of "medium" width that lies within the VNIR portion of the spectrum. Panchromatic sensors comprise only one channel whose SRF spans over a broader wavelength range, also within the VNIR portion. The reason for the choice of this spectral region is twofold: On the one hand, the visible range is where the human eye is sensitive in, which allows for the generation of easily interpretable "true-color RGB" images; On the other hand, this portion is particularly suitable for passive optical satellite remote sensing, because this is where both the Sun has its peak irradiance at the top of the atmosphere and the atmosphere is lowly opaque (compare Fig. 3). Moreover, since instruments with broader SRFs can detect more energy, by means of photons, they are capable of generating higher-spatial resolution imagery at a similar SNR. This trade-off between spectral and spatial resolution has been used in many satellite missions such as WorldView-1 through 4, QuickBird and the SPOT series, which carry both a multispectral sensor for spectral information and a panchromatic sensor for higher-spatial resolution imagery. Fig. 13 depicts the SRFs of the QuickBird (left) and WorldView-2 (right) satellite missions, representatively for such a multispectral-panchromatic combination. The fusion of a multispectral image with a higher-resolution panchromatic image is one of the focusses of this thesis and will be introduced in Section 2.2.1.

As the technological progress and research continuously leads to advancements in sensor design, state of the art commercial satellite-based multispectral instruments achieve a spatial resolution of around 1 m GSD at best nowadays; The GSD of panchromatic sensors can be even well below 1 m GSD. The WorldView-4 multispectral and panchromatic instruments, for instance, provide up to 1.24 m and 0.31 m GSD imagery, respectively. More missions and their specifications are listed in Table 1 on page 38.

The high to very high spatial resolution of multispectral and panchromatic instruments both characterises and drives the user applications with this type of data. Apart from the obvious facilitation of the visual interpretability by the human eye through higher spatial resolution, the majority of applications utilize the perceptibility of *object shapes*, *surface structures* and even *textures*. Figures 14 through 17 illustrate examples of recent applications which explicitly utilize the high or very high resolution of multispectral and panchromatic imagery. Bar and Raboy [2013], for instance, developed an automatic moving car detection algorithm,

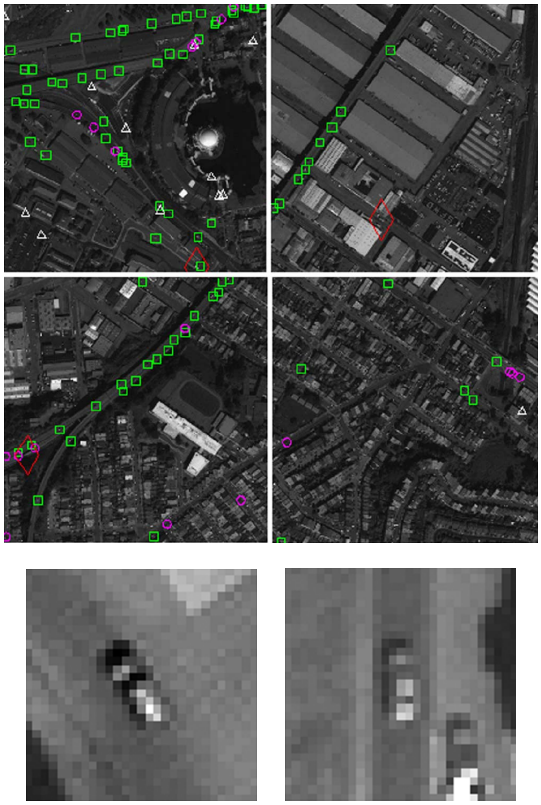


Fig. 14. Moving car detection application to a WorldView-2 multispectral+panchromatic image pair. Upper figure: Correctly (green), falsely (purple) and undetected (red) vehicles. Lower figure: Magnification of two cars. *(Developed and published by Bar and Raboy [2013]; All credits for this figure belong to Bar and Raboy)*

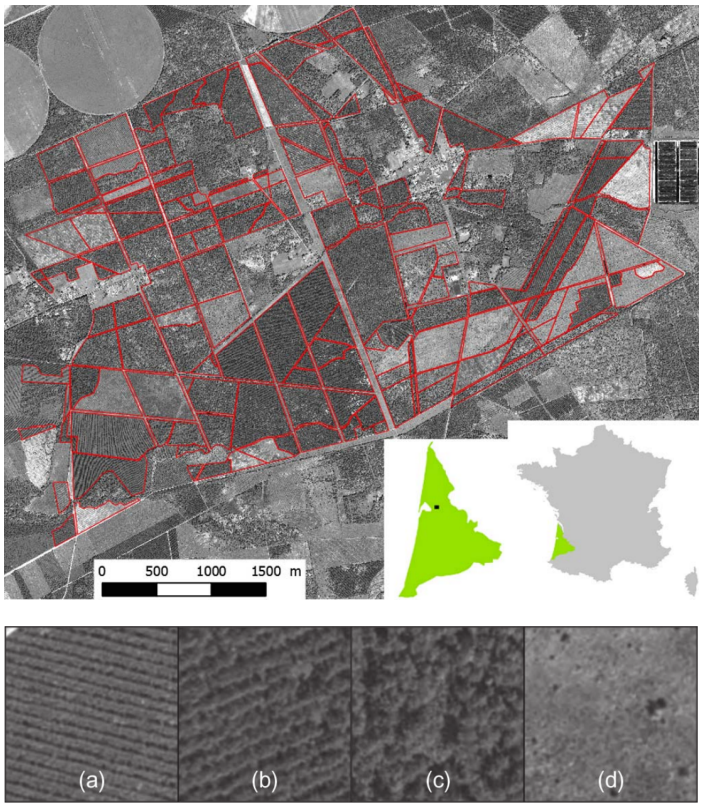


Fig. 15. Wavelet-based texture feature extraction for the classification of forest age classes using 0.5-m GSD panchromatic Pléiades imagery. Upper figure: Contours of separate forest stand regions. Lower figure: Representative forest age classes: (a) 1-9 years; (b) 10-19 years; (c) >19 years; (d) Clear cuts. *(Developed and published by Regniers et al. [2015]; All credits for this figure belong to Regniers et al.)*

which takes into account the short time difference between the panchromatic and the multispectral data acquisitions (Fig. 14). Regniers et al. [2015] developed a wavelet-based texture feature extraction method for the classification of forest age classes and applied it to 0.5-m GSD panchromatic Pléiades imagery (Fig. 15). A more general object detection method based on rotation-invariant convolutional neural networks was developed by Cheng et al. [2016] and applied to a large collection of high- and very high-resolution imagery. In a similar manner, Shu et al. [2015] proposed a spatial object-based unsupervised classification algorithm for very high-resolution panchromatic satellite imagery. Furthermore, high-spatial resolution panchromatic and multispectral imagery is well combinable with high-spatial resolution data acquired by a system that has a similar viewing geometry as the multispectral or panchromatic sensor at hand. Similarly to most passive optical remote sensing systems, Earth observing LiDAR systems, operated from an airplane or satellite platform, are typically nadir- or nearly nadir-looking [Popescu et al., 2011; Vierling et al., 2013]. Hermosilla et al. [2011] developed an automatic building detection algorithm based on combined high-resolution multispectral and LiDAR data (Fig. 17).

As can be reformulated for any of the introduced types of resolution, very high spatial resolution generally entails high redundancy in the data wherever the distinguishable objects on the ground are in a spatial scale larger than the sensor's GSD. This is why data compression in the spatial domain, and spatial transformations in general, are increasingly important in the processing chains of current an future multispectral and panchromatic satellite mission.

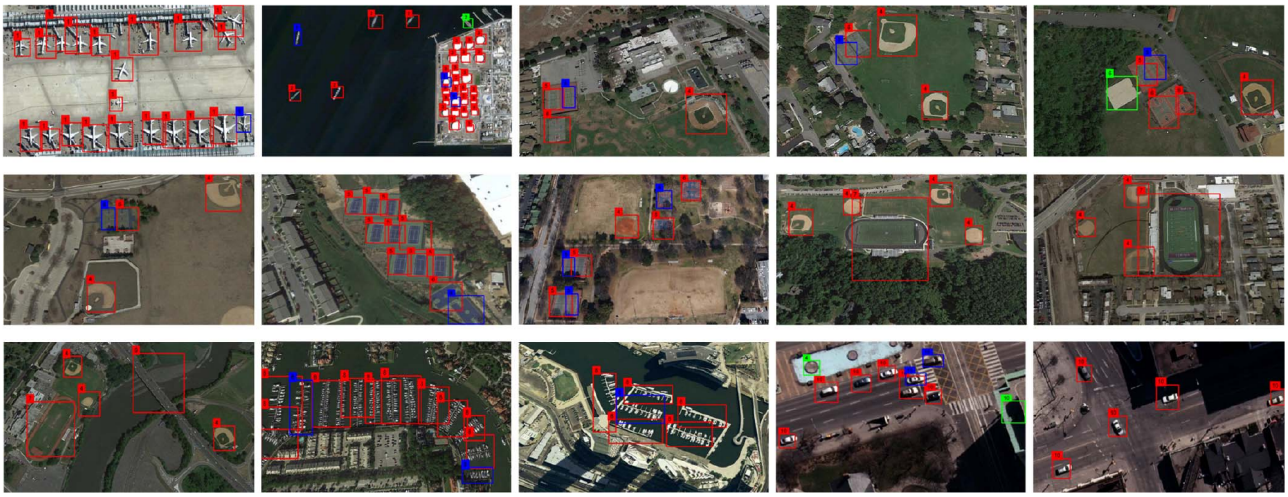


Fig. 16. Object detection in very high-resolution multispectral imagery based on rotation-invariant convolutional neural networks. *(Developed and published by Cheng et al. [2016]; All credits for this figure belong to Cheng et al.)*

Fig. 17. Automatic building detection approach, combining high-resolution multispectral imagery and LiDAR data. *(Developed and published by Hermosilla et al. [2011]; All credits for this figure belong to Hermosilla et al.)*

In addition to the increasingly rich spatial information content provided by multispectral imagers, the multiband characteristics allow for applications beyond visual RGB true-color image inspection. Spectral transformations and band ratios give rise about specific spectral features of materials on the ground and allows for material discrimination to some extend. One of the widely used band ratios in multispectral remote sensing is the normalized difference vegetation index (NDVI) which – similarly to other band ratios that involve both a VNIR and a red channel – is used to analyse an image for green vegetation. Band ratios are used to generate an image pixel-wise from the multispectral image, in which specific materials (or material classes) in the image are *illuminate* while anything else is darkened.

To conclude this section, multispectral and panchromatic sensors provide imagery of increasingly high spatial resolution. Such data is particularly valuable for observations of areas or objects, where the level of spatial details is in the range of the sensor's GSD or finer. Recent developments in shape- and structure-based object detection for the detection of cars, buildings, airplanes, ships and others, indicate that urban mapping and applications involving man-made structures are where the potential and value of this (very) high-spatial resolution imagery lies. Moreover, multispectral sensors have, to some extent, material discrimination capabilities, which can be accentuated by applying a suitable band ratio.

As illustrated in Fig. 10 on page 16, materials can be potentially more accurately discriminated by a hyperspectral sensor, which leads us to the following section.

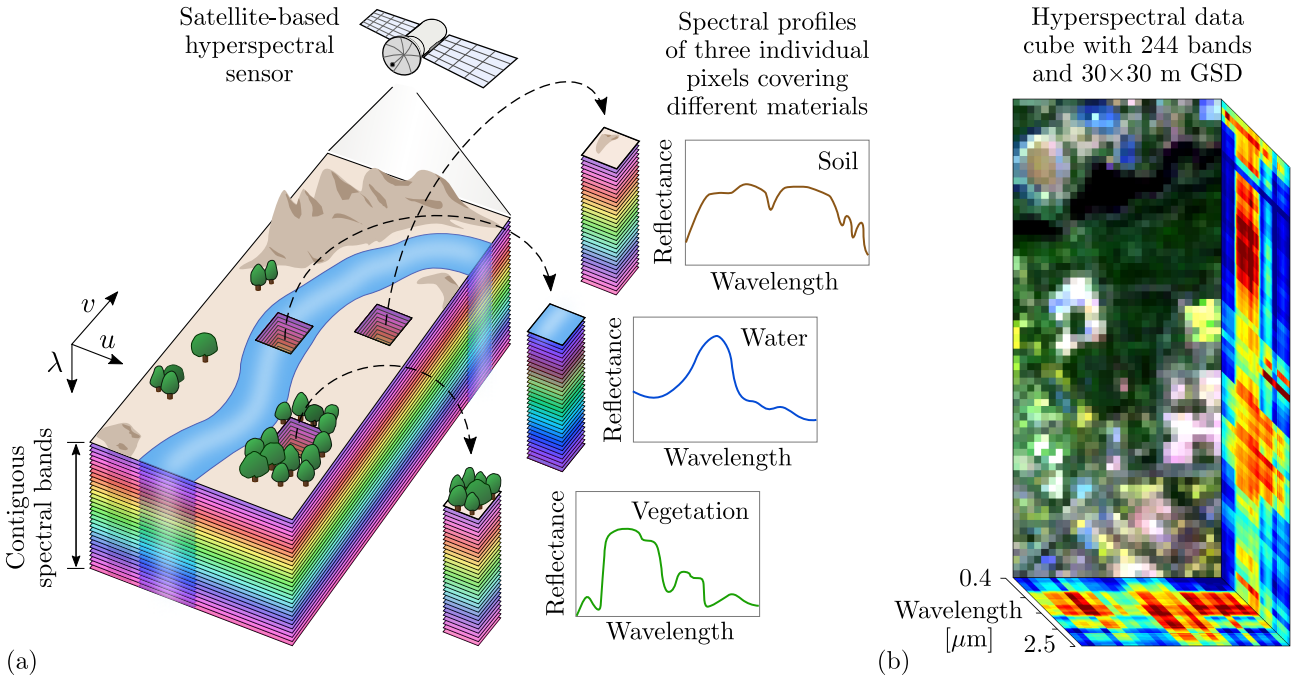


Fig. 18. (a) Concept of spaceborne hyperspectral remote sensing. A satellite-based sensor acquires information of an area on the ground simultaneously in a large number of contiguous spectral channels. Typical scanning concepts include the above-described whiskbroom and pushbroom techniques (see Fig. 6), where the latter method is more commonly applied in hyperspectral remote sensing. Since the rich spectral information allows for material identification, the data is usually processed up to the reflectance level in order to facilitate the comparability of pixel spectral profiles to known material spectral signatures. The three spectral profiles illustrate typical reflectance curves for soil, water and vegetation. *(Graphics inspired by Shaw and Burke [2003, p. 4])*. (b) Real hyperspectral data acquired by the HySpex airborne sensing system over Munich in 2012. The data shown is a spectrally and spatially degraded and down-sampled version of the original very high-resolution HySpex data, to match the sensor characteristics of the satellite-based EnMAP instrument.

2.1.5 Specifics in Hyperspectral Remote Sensing

Hyperspectral Data Hyperspectral data is characterized by a large number (dozens to hundreds) of mostly contiguous spectral channels each of which is responsive in a relatively narrow wavelength range within the reflective – and possibly emissive – optical portion of the spectrum. As briefly discussed on pages 15 to 18, a large number of bands usually implicates advanced material discrimination capabilities and, hence, a high spectral resolution. While multispectral data has a significantly larger number of pixels in both spatial directions than number of bands, it is commonly referred to as a multispectral *image*. Hyperspectral data, on the other hand, not only comprises more spectral bands than multispectral images, but, due to the typically larger GSDs and narrower swaths of hyperspectral sensors, also contains fewer pixels in the spatial directions. Thus, hyperspectral data is commonly represented and illustrated as a 3-D cube with two spatial dimensions in along- and across-track directions, and the spectral dimension built up by the bands. Fig. 18 depicts this cubical representation, as well as the connection between hyperspectral pixels on the Earth's surface, corresponding materials, and spectral profiles. Moreover, as mentioned in the introduction of spectral resolution on page 17, most spectral characteristics of a sensor can be well explained via the sensor's SRFs. An illustration of representative hyperspectral SRFs in comparison to SRFs of various multispectral sensors can be found in Fig 10 on page 16.

Calibration The basic concept of instrument and data calibration and corresponding processing steps for hyperspectral instruments are principally identical to those of multispectral sensors. However, some calibration steps are more challenging for hyperspectral sensors in particular, due to the following reasons:

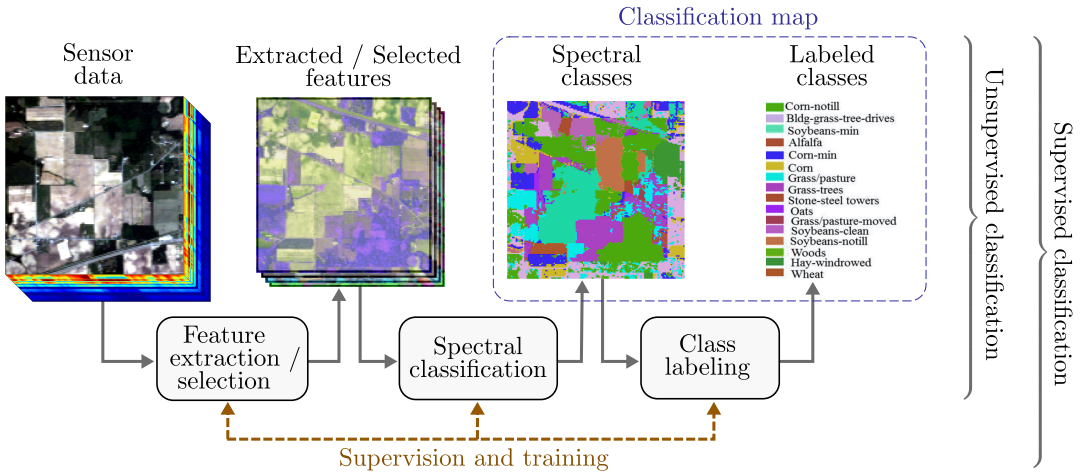


Fig. 19. General flowchart and components of hyperspectral data classification. The underlying hyperspectral data used for illustration are the freely available and widely used "Indian Pines" data acquired by the airborne AVIRIS instrument. *(The classification map image including labels are borrowed from Huang et al. [2014] and the flowchart structure was partially designed after Eismann [2012, p. 564])*

- The large number of spectral bands implicate a large number of FPA elements which require individual calibration.
- Some of the narrow contiguous wavelength ranges, in which the hyperspectral bands are responsive, coincide with atmospheric absorption features. Corresponding bands are differently affected by the atmosphere than – possibly neighboring – channels that do not coincide with these features. Bands falling into low atmospheric transmittance regions measure weaker signals and, therefore, have lower SNRs. This varying influence of the atmosphere must be carefully taken into account in the atmospheric correction procedure in order for the hyperspectral data to preserve its discrimination capability of material absorption features.
- For hyperspectral sensing systems, the *sensor* calibration phase is particularly critical. One reason for that is the so-called *smile* effect, which is a spectrum shift varying with the across-track (in the case of pushbroom scanners) detector number. Caused by aberrations in the imaging optics it leads to the visual effect of the image looking "curved". For space-based sensors this effect requires compensation only once during the initial phase of the mission and does not change notably over the sensor's lifetime [Schowengerdt, 2006, pp. 341-343].

Classification One of the primary objectives of hyperspectral data acquisition and processing is to find, identify and discriminate materials whose spectral signatures are hardly, or not at all, differentiable in multispectral data. The ideal product for many hyperspectral applications is a *classification map* indicating which materials are present at what location in the scene. Giving the importance of this classification task and the diversity of applications which is growing with the announcement of planned hyperspectral satellite missions (see Section 2.1.7), developing generally applicable, as well as application-specific classifiers remains an active research field [Ghamisi et al., 2016; Li et al., 2016; Aptoula et al., 2016; Gurram et al., 2016; Zhong et al., 2015]. Fig. 19 depicts a general flowchart of hyperspectral data classification whose individual components are described in the following.

Sensor data can be available in multiple formats corresponding to different levels of calibration and atmospheric compensation such as DNs, radiance or reflectance. If there is prior information available on the spectral signatures corresponding to the materials of interest, those signatures are usually given in reflectance and belong to spectral libraries. If such information is accounted for in the classification algorithm design, the sensor data should

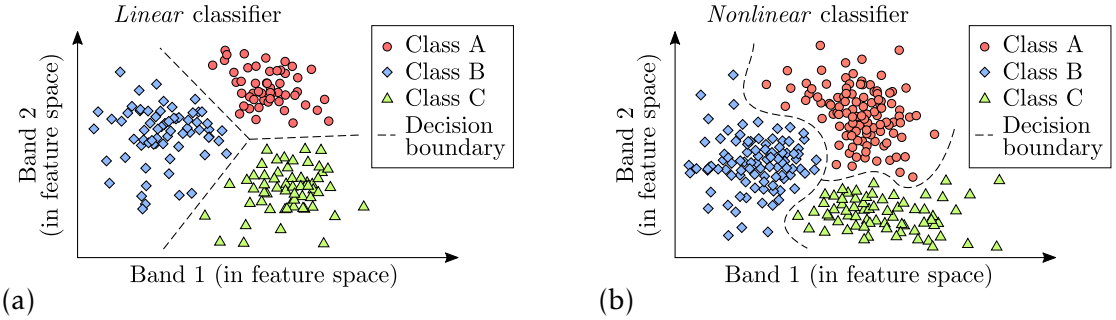


Fig. 20. Illustration of a linear (a) and nonlinear (b) classifier in 2-D scatter plots. Both examples show pixel values in one feature band plotted over the same pixel's value in another feature band. Given the point distribution and possibly additional prior knowledge, the points can be clustered and assigned to one of the spectral classes of interest according to discrimination boundaries which are defined by the underlying the classification algorithm. While linear classifiers define sets of hyperplanes (a), nonlinear classifiers define nonlinear submanifolds (b). In practice, depending on the data and application, the number of materials is usually much larger and the dimension of the feature space much higher than illustrated in this figure.

be given in surface reflectance as well. Many classification approaches, however, are invariant to the level of calibration [Eismann, 2012, p. 563]. Given the sensor data at a suitable processing level, the main classification steps can be broadly described as folows:

Feature extraction/selection: Transformation of the data onto a lower-dimensional subspace, in which the information of interest is captured and represented more efficiently, compact and possibly less noisy. Apart from saving memory and facilitating computational efficiency, feature space transformations can significantly improve the classification accuracy [Zhang et al., 2016; Li et al., 2014]. Feature selection represents a special case of feature extraction, in which a subset of all available bands in the data is *selected*. More general feature extraction strategies involve spectral transformations some of which are introduced in Section 2.1.6 below.

Spectral classification: The actual process in which each pixel, i.e. spectral profile, is assigned to a specific spectral class usually corresponding to a specific material. This assignment is the more challenging, the higher the spectral similarity among the materials of interest. This similarity is generally higher for large numbers of materials and if the materials of interest belong to the same categorie, such as different types of trees. The classification task is conducted in a suitable feature space and requires the development of a *classifier* or *discrimination function* capable of separating the pixels corresponding to different classes. Such a classifier can be linear or nonlinear as illustrated in Fig. 20.

Class labeling: Association of each spectral class with a physical identifier usually corresponding to a material or category of materials.

If there is no ground reference data, training samples or other means of prior knowledge on the correspondence between the given data and the materials of interest, all steps described above must be performed autonomously. In this case, the classification process is called *unsupervised*. If, in contrast, any of the afore-mentioned information exists and is implemented in the classification algorithm design, the process is called *supervised*. Manually labeled sample spectra and ground reference data are particularly useful to achieve accuracy improvements over purely unsupervised classification approaches [Condessa et al., 2016; Sun et al., 2015].

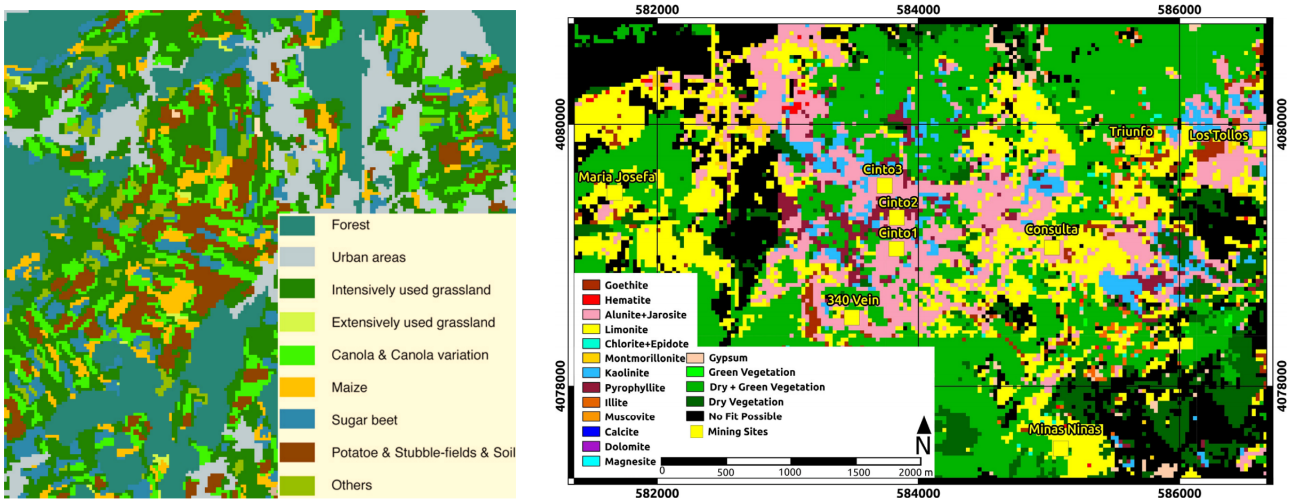
The development of classification algorithms has been a vast topic not only in the hyperspectral community, but even more so in the general field of machine learning and many of its branches, where the underlying theories, the classifier designs are based upon, are as vast and divers as their applications. Well-established categories of (supervised) classification algorithms include *Support Vector Machines* (SVMs) [Schölkopf et al., 1999; Boser et al.,

1992; Waske et al., 2010] and *Random Forests* (RFs) [Liaw and Wiener, 2001; Svetnik et al., 2003; Ristin et al., 2016], to name only two of many.

Spectral Unmixing In reality, the Earth's surface is highly heterogeneous. Depending on the imaged scene, the level of details and diversity of materials can differ significantly. On the other hand, hyperspectral sensors – especially space-based ones – are limited in spatial resolution as explained above. Consequently, occurrences of multiple material constituents within the same pixel are inevitable. Especially in urban areas, pure pixels, i.e. pixels containing only one material, are rare. Spectral unmixing is the process of (1) *identifying endmembers*, i.e. spectral signatures corresponding to the materials present in each pixel, and (2) their *abundances*, i.e. relative contribution of each endmember to the pixel's spectral profile [Keshava and Mustard, 2002; Chang, 2003]. The process of *how* the energy is mixed on the ground to form the resulting measured spectral profile, depends on the materials present in the pixel as well as the 3-D structure and topology of the observed scene. Since these factors are generally unknown, the mathematical model most commonly used to describe the mixing process is the *linear mixing model* (LMM) [Manolakis et al., 2001; Haertel and Shimabukuro, 2005]. However, since the LMM is based on the strong assumption of single scattering processes on the ground, there has been much development recently towards more sophisticated non-linear models that account for information about specific observed materials, 3-D structures and implicated physical light scattering processes [Heylen and Gader, 2014; Févotte and Dobigeon, 2015; Heylen and Scheunders, 2016].

Applications Hyperspectral remote sensing has a wide array of applications. With an increasing availability of portable imaging spectrometers, hyperspectral remote sensing has been gaining popularity beyond the Earth observing remote sensing community. To date, terrestrial hyperspectral imaging applications include microscopic medical [Bellisola and Sorio, 2011], biological [Gosnell et al., 2016] and chemical [Dill et al., 2016; Gowen et al., 2011] imaging, historical manuscript research [Joo Kim et al., 2011], art [Cucci et al., 2016], archaeology [Vandenabeele and Donais, 2016] and food processing [Dai et al., 2014], to name only a few. Furthermore, there is a growing market of UAV-based and airborne hyperspectral systems [Proctor and He, 2015; Köhler, 2016; Vreys et al., 2016], and multiple satellite-based Earth observing hyperspectral instruments will be launched in the near future (see Section 2.1.7). This development has led to a growth also in Earth observation applications. Established and emerging fields include mineralogy [Schneider et al., 2014; Ting-ting and Fei, 2012; Kruse et al., 2003], agriculture [Teke et al., 2013], geology [Murphy et al., 2012], surveillance [Polo et al., 2015; Yuen and Richardson, 2010], detection of oil seeps [Ellis et al., 2001], forestry [Başkurt et al., 2015; Goodenough et al., 2012; White et al., 2010], invasive species [Walsh et al., 2008], canopy chemistry [Wu et al., 2010; Asner, 2008; Martin et al., 2008], disturbances [Numata et al., 2010; Mitri and Gitas, 2010], volcanism and geology [Abrams et al., 2013; Davies et al., 2006], water resources and quality [Gursoy et al., 2015; Fan, 2014; Wu et al., 2014; Govender et al., 2007], target detection [Nasrabadi, 2014; Yang and Shi, 2016], change detection [Wu et al., 2013], climate change [Shobiga and Selvakumar, 2015], natural hazard assessment and monitoring [Bonis et al., 2015; Zhang et al., 2011], as well as mapping and monitoring of urban areas [Hardin and Hardin, 2013; Shafri and Taherzadeh, 2012; Marino et al., 2001]. Fig. 21 depicts classifications maps corresponding to (a) agricultural and (b) to mineral mapping applications based on space-based HYPERION and simulateed EnMAP data, respectively.

Considering this diverse potential of hyperspectral data utilization, it is unfortunate that only a few of the above-listed applications are directly applicable to *satellite-based* hyperspectral data. This is partly because of the data's coarse spatial resolution and the consequent mixture of materials and objects in the pixels. On the other hand, satellites are best



(a) (b)

Fig. 21. (a) Land use classification map derived from EO-1 HYPERION data acquired over the Limpach Valley on August 18, 2002 *(All credits for this figure belong to Eckert and Kneubühler [2004])*. (b) Material map of the Rodalquilar gold-alunite deposit in Spain, developed from HyMAP-based simulated EnMAP data *(All credits of this figure belong to Mielke et al. [2016])*.

suitable for large-scale and global mapping and monitoring applications. Considering also the aspects of data storage and transfer, it would be desirable to have an improved spatial resolution specifically in those areas where higher levels of details can be expected or there is a particular user interest, such as urban areas, specific agricultural regions, mineral mining areas, etc. Larger homogeneous areas, in contrast, may not require higher spatial resolution or data acquisition in the first place.

One part of this work aims at enhancing the spatial resolution of hyperspectral data in a selective or systematic manner to open up new fields of application by combining the best of satellite-based (e.g. global coverage and access, continuous operation and flexibility to provide service in urgent matters) and high-resolution (e.g. high fidelity and accuracy in spatial material discrimination) hyperspectral remote sensing capacities.

2.1.6 Data Representation, Statistical Parameters and Spectral Transformations

Data Representation As mentioned in the previous sections, optical remote sensing data has two spatial dimensions and one spectral dimension, where, in the panchromatic case, the spectral dimension simply equals one. Let $\boldsymbol{X}_{3\mathrm{D}} \in \mathbb{R}^{U\times V\times K}$ denote an arbitrary remote sensing image with U and V pixels in along- and across-track direction, respectively, and K spectral bands [5]. Especially if $K > 1$, $\boldsymbol{X}_{3D}$ contains spatial and spectral information which should, optimally, both be exploited in analysis and processing procedures. In hyperspectral data processing in particular, the relatively low spatial and high spectral resolution makes the spectral information content the primary driving factor in data algorithm development. In order to facilitate mathematical modelling and data analysis not only for hyperspectral, but also for multispectral imagery, it can be of avail to combine the two spatial dimensions,

[5] In fact, after geometric correction and, possibly, registration to a different coordinate system, real 3-D remote sensing data does not necessarily have a rectangular spatial extend. However, the data can always be padded with zeros, which is why this representation can be used without loss of generality.

U and V, to a single one, and work with the following 2-D matrix representation instead:

$$\boldsymbol{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N] \tag{22}$$

$$= [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_K]^T, \tag{23}$$

where each column vector

$$\mathbf{x}_i = [x_{1,i}, x_{2,i}, \ldots, x_{K,i}]^T \tag{24}$$

denotes a spectral profile corresponding to the spatial pixel $i \in \{1, \ldots, N\}$ and $N = U \cdot V$ is the *total* number of pixels in the data. Furthermore, the row vectors $\boldsymbol{x}_k$, $1 \leq k \leq K$, in Eq. (23)

$$\boldsymbol{x}_k = [x_{k,1}, x_{k,2}, \ldots, x_{k,N}]^T \tag{25}$$

represent vectorized image bands, where $x_{k,i}$ denotes the data value at pixel i in band k whose SRF is centered at wavelength λ_k. Depending on the processing level, $x_{k,i}$ can be a given as DN, radiance or reflectance value:

$$x_{k,i} = \begin{cases} DN_i(\lambda_k) & \text{if } \boldsymbol{X}_{3\mathrm{D}} \text{ is given in digital numbers} \\ L_i(\lambda_k) & \text{if } \boldsymbol{X}_{3\mathrm{D}} \text{ is given in at-sensor radiance values} \\ \rho_i(\lambda_k) & \text{if } \boldsymbol{X}_{3\mathrm{D}} \text{ is given in surface reflectance values} \end{cases} \tag{26}$$

Note that the notation used in Eq. (26) is chosen in accordance with Eq. (4) and (17). Importantly, the two spatial dimensions are implicitly maintained in the linear rastered ordering in Eq. (22). Thus, it remains possible to perform 2-D spatial operations.

Statistical Parameters The *(sample) mean* and *(sample) standard deviation* of $\boldsymbol{X}$ in band k can be calculated by

$$\overline{\boldsymbol{x}}_k = \frac{1}{N} \sum_{i=1}^{N} x_{k,i} \tag{27}$$

and

$$\sigma_{x_k} = \left(\frac{1}{N+1} \sum_{i=1}^{N} (x_{k,i} - \overline{\boldsymbol{x}}_k)^2 \right)^{1/2}, \tag{28}$$

respectively. Furthermore, the *covariance matrix* $\boldsymbol{C}$ and *covariance* σ_{x_k,x_l} between bands k and l are respectively defined as

$$\boldsymbol{C} = \left[\sigma_{x_k,x_l} \right]_{k,l=1}^{K} := \begin{bmatrix} \sigma_{x_1,x_1} & \cdots & \sigma_{x_1,x_K} \\ \vdots & \ddots & \vdots \\ \sigma_{x_K,x_1} & \cdots & \sigma_{x_K,x_K} \end{bmatrix} \tag{29}$$

and

$$\sigma_{x_k,x_l} = \frac{1}{N+1} \sum_{i=1}^{N} (x_{k,i} - \overline{\boldsymbol{x}}_k)(x_{l,i} - \overline{\boldsymbol{x}}_l). \tag{30}$$

Similarly, $\boldsymbol{\mathcal{R}}$ and $r_{k,l}$ denote the *correlation matrix* and the *correlation coefficient* between bands k and l, and are respectively defined as

$$\boldsymbol{\mathcal{R}} = [r_{k,l}]_{k,l=1}^{K} \tag{31}$$

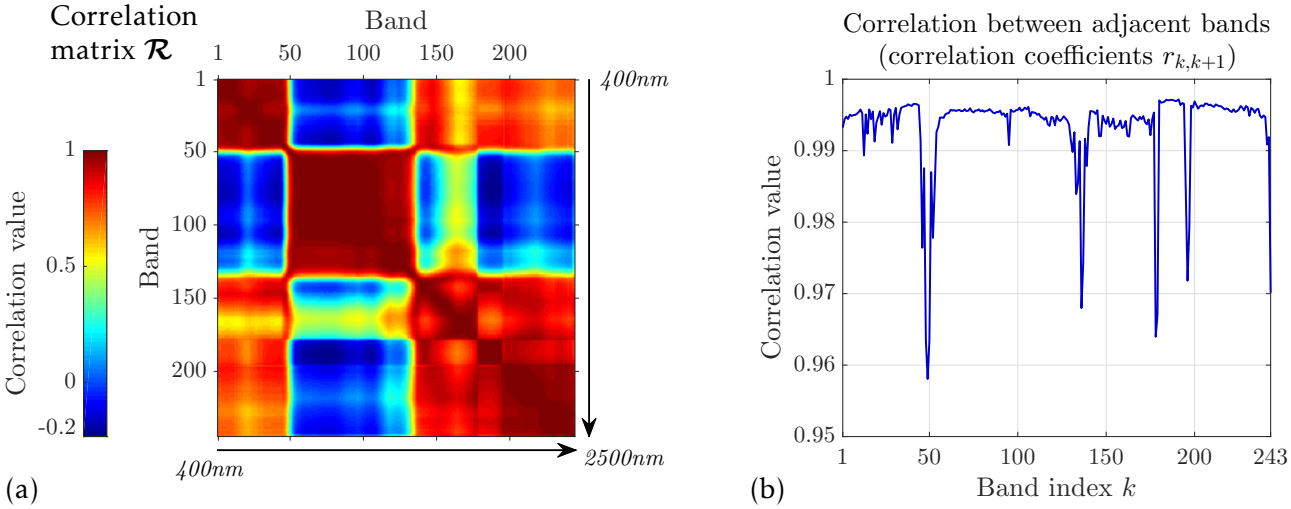


Fig. 22. (a) Correlation matrix calculated from the simulated 244-band EnMAP data shown in Fig. 18 (b). Noticeable are the square blocks along the diagonal colored in dark red, which indicate high mutual correlation among bands in groups corresponding to connected wavelength ranges. (b) Correlation coefficients $r_{k,k+1}$ of contiguous bands. This graph directly relates to Fig. (a) in that it plots any of the two next-to-diagonal slices of the correlation matrix. Note that within the afore-mentioned groups associated with the dark-red blocks in Fig. (a), the correlation between adjacent bands exceeds 0.99 for nearly all band indeces k.

and

$$r_{k,l} := CC(x_k, x_l) = \sigma_{x_k,x_l}/(\sigma_{x_k}\sigma_{x_l}) \quad (-1 \leq r_{k,l} \leq 1). \tag{32}$$

Fig. 22 illustrates the correlation matrix $\mathcal{R}$ (left) and the correlation coefficients $r_{k,k+1}$ of adjacent channels (right), calculated from the simulated 244-band hyperspectral EnMAP data shown in Fig. 18 (b). Noticeable are the square blocks along the diagonal colored in dark red, which indicate high mutual correlation among bands in groups corresponding to connected wavelength ranges. The figure on the right hand side directly relates to the one on the left hand side in that the former plots any of the two next-to-diagonal slices of $\mathcal{R}$. Note that within the afore-mentioned groups associated with the dark-red blocks in Fig. (a), the correlation between adjacent bands exceeds the high value of 0.99 for nearly all band indeces k.

Spectral Transformations and Dimensionality Reduction For many applications and analyses of multiband (hyperspectral or multispectral) data, it is advantageous to transform $\boldsymbol{X}$ from the original image space to some feature space. Such a transformation can be generally written as

$$\boldsymbol{X}' = f(\boldsymbol{X}) \tag{33}$$

with f being some transformation function. Although transformations do not add new information to the original data, they may reveal, redistribute and concentrate information, which is contained and possibly spread or hidden in the original data, for an easier access and analysis of the features of interest. If f is *linear*, the data representation inherent in $\boldsymbol{X}$ allows for explicit *spectral* and *spatial* transformations via left- and right-sided multiplication with $\boldsymbol{X}$, respectively:

$$\boldsymbol{X}' = \boldsymbol{W}_{\text{spectral}} \cdot \boldsymbol{X} + \boldsymbol{B} \qquad \textit{(spectral transformation)} \tag{34}$$

$$\boldsymbol{X}' = \boldsymbol{X} \cdot \boldsymbol{W}_{\text{spatial}} + \boldsymbol{B} \qquad \textit{(spatial transformation)} \tag{35}$$

Both spectral and spatial transformations, as given by Eq. (34) and (35), will be frequently used in the following chapters. While the spatial transformations relevant for this thesis mainly include filtering of $\boldsymbol{X}$ with PSFs and subsequent down-sampling operations, the

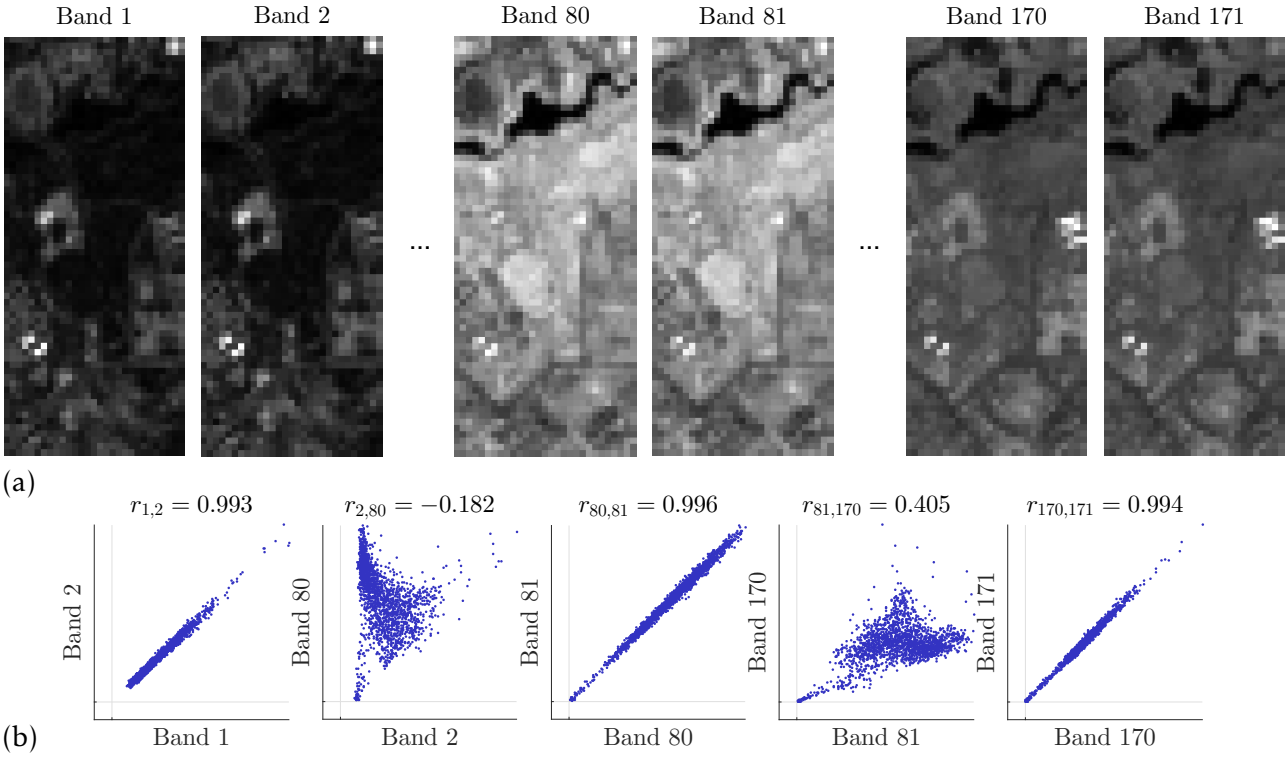


Fig. 23. (a) Selected gray-scaled band images of simulated EnMAP data shown in Fig. 18 (b). High correlation between adjacent bands can be visually observed. Noticeably, the wavy water area in the top part of the image features low intensity values in all three wavelength groups, while the majority of the remaining pixels correspond to (green) vegetation which are brightest around the NIR bands 80/81. (b) Scatter plots of intensity image band pairs illustrate significant correlation differences between contiguous and distant band pairs. This difference is quantitatively confirmed by the corresponding correlation coefficients displayed above the scatter plots.

spectral transformations used in Section 5 should be formally introduced, which is what the remainder of this subsection focuses on.

Multispectral and, even more so, hyperspectral image bands are often highly correlated as we have seen in Fig. 22. That is to say, bands are numerically as well as visually similar. Fig. 23 illustrates this similarity for three pairs of adjacent bands, by means of selected gray-scaled band images and corresponding scatter-plots.

Correlation between bands is highly material-dependent. A homogeneous scene containing only green vegetation, for instance, would lead to mutually correlated bands within those wavelength ranges in which the vegetation spectral signatures do not change much, i.e. from 400 to 700 nm and from 850 to 1300 nm. Obviously, high mutual correlation implicates *redundancy* in the data. Furthermore, the higher the redundancy, the less independent information is hold in the data. The amount of independent information can be quantified by the *intrinsic spectral dimension*, $K_{\text{sub}} \leq K$, of the data.

Spectral transformations can be used to project the original data onto a basis with respect to which the information of interest is concentrated and separated from the less relevant portion of the data. The simplest, yet not linear, spectral transformations are band ratios including vegetation – or other material – indices such as the above-mentioned NDVI. They are more commonly applied to multispectral data and the information of interest can be interpreted as a pixel-wise likelihood of the presents of a specific type of material. The linear transformations that are more relevant for this work redistribute the information contained in the data linearly from most to least significant. The less informative part of the data is then considered redundant and is, therefore, commonly discarded. K_{sub} is usually unknown

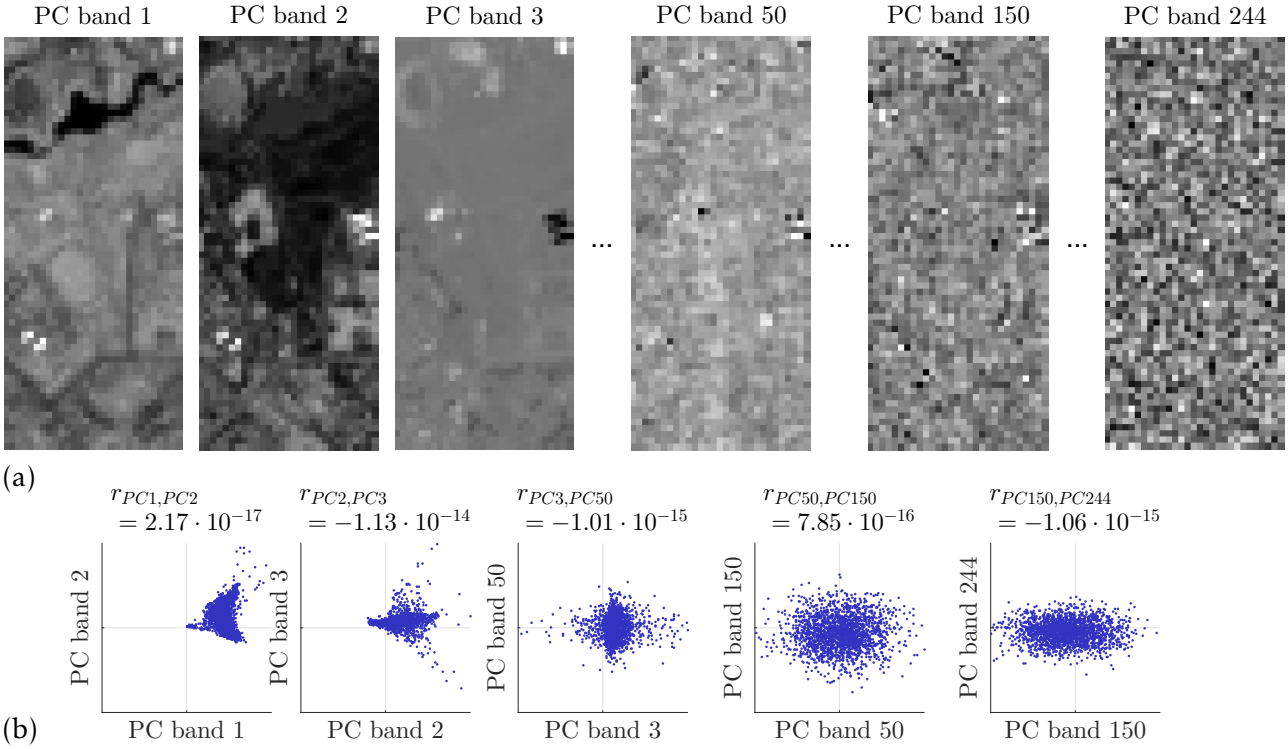


Fig. 24. (a) Selected gray-scaled band images of PC-transformed simulated EnMAP data. Each of the leading PC bands show different spatial structural details. Observably, both meaningful information and SNR visually decrease with increasing PC band index. (b) Scatter plots corresponding to the intensity images shown in Fig. (a). Compared to the scatter plots in Fig. 23 (b), the points in all plots shown here are distributed rather equally around one of the two main axes, thus indicating very low correlation.

and therefore estimated from the data as will be discussed below. Once K_{sub} is estimated, only the most meaningful K_{sub} "feature bands" in the projected data $\boldsymbol{X}'$ are kept:

$$\boldsymbol{X}' = \begin{bmatrix} \boldsymbol{X}'_{\text{keep}} \\ \boldsymbol{X}'_{\text{discard}} \end{bmatrix} = \left[\begin{bmatrix} x'_{1,1} & \cdots & x'_{1,N} \\ \vdots & & \vdots \\ x'_{K_{\text{sub}},1} & \cdots & x'_{K_{\text{sub}},N} \end{bmatrix} \begin{bmatrix} x'_{K_{\text{sub}}+1,1} & \cdots & x'_{K_{\text{sub}}+1,N} \\ \vdots & & \vdots \\ x'_{K,1} & \cdots & x'_{K,N} \end{bmatrix} \right] \downarrow \text{Decreasingly "significant" information} \tag{36}$$

Traditional linear transformations used in the spectral domain include, but are not limited to, *Principal-Component Analysis* (PCA) [Pearson, 1901] (also known as *Karhunen-Loewe* (KL) [Karhunen, 1947] or *Hotelling* transformation [Hotelling, 1933]), *Singular Value Decomposition* (SVD) [Stewart, 1993] and *Maximum Noise Fraction* (MNF) [Green et al., 1988] (also referred to as *Minimum Noise Fraction* [Datt et al., 2003] and *Noise-Adjusted Principal Components* (NAPC) transform [Lee et al., 1990]), all of which are strongly interconnected. In the following, only PCA will be briefly introduced and brought into context.

PCA is a linear feature space transformation of the type shown in Eq. (34) with $\boldsymbol{B} = \boldsymbol{0}$:

$$\boldsymbol{X}_{\text{PC}} = \boldsymbol{W}_{\text{PC}} \cdot \boldsymbol{X} \tag{37}$$

$\boldsymbol{W}_{\text{PC}}$ is of particular interest in hyperspectral remote sensing, as it diagonalizes the data

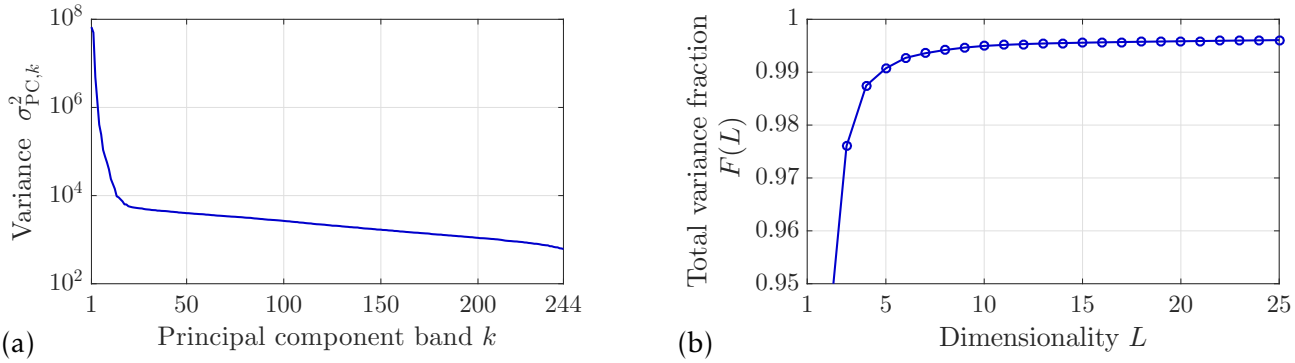


Fig. 25. (a) Variance of PC-transformed simulated EnMAP data in decreasing ordering. (b) Total variance fraction calculated from the variance values shown in Fig. (a). Noticeable, the leading five PC bands capture over 99 % and the leading 10 bands over 99.5 % of the total variance.

covariance matrix $\boldsymbol{C}$ and, thus, removes the inter-band correlation:

$$\boldsymbol{C}_{\mathrm{PC}} = \boldsymbol{W}_{\mathrm{PC}} \boldsymbol{C} \boldsymbol{W}_{\mathrm{PC}}^T = \mathrm{diag}\left([{\sigma_{\mathrm{PC},1}}^2, \ldots, {\sigma_{\mathrm{PC},K}}^2] \right) \tag{38}$$

In Eq. (38), $\sigma^2_{\mathrm{PC},k}$ the represents variance of the k-th principal component (PC) band, i.e. row in $\boldsymbol{X}_{\mathrm{PC}}$, and can be found as the k-th largest eigenvalue of $\boldsymbol{C}$. Having $\sigma^2_{\mathrm{PC},k}$, the k-th row of the PC transformation matrix $\boldsymbol{W}_{\mathrm{PC}}$ is determined as the corresponding unit eigenvector $\boldsymbol{e}_k$:

$$\boldsymbol{W}_{\mathrm{PC}} = [\boldsymbol{e}_1, \ldots, \boldsymbol{e}_K]^T \tag{39}$$

Hence, $\boldsymbol{W}_{\mathrm{PC}}$ rotates $\boldsymbol{X}$ into an orthogonal basis whose axes are called *principal directions*.

Fig. 24 shows selected bands of the PC-transformed simulated EnMAP data, allowing for a direct comparison to the original data shown in Fig. 23. Comparing the scatter plots in both figures reveals the PC rotation property into the main axes. Furthermore, the displayed nearly-zero correlation coefficient values confirm the aforementioned correlation-removal property of PCA. Moreover, the SNR visually decreases as the PC band index increases. This phenomenon is directly related to the monotonically decreasing variance, which is depicted in Fig. 25 (a). This variance curve has a typical shape in that there is an x-value (band number) from which on the variance "flattens". This data-specific point usually determines K_{sub}, which, according to the linear mixing model, may correspond to the unknown number of materials in the observed scene [Haertel and Shimabukuro, 2005].

Determining K_{sub} and, hence, the inherent spectral subspace dimensionality, is an ongoing research topic [Rasti et al., 2015a; Benner et al., 2015; Torti et al., 2014; Bioucas-Dias and Nascimento, 2008]. One rather intuitive method selects K_{sub} such that the leading subspace captures a sufficiently high fraction of the total data variance, which, for any dimension $1 < L < K$, is defined as [Eismann, 2012, pp. 526-529]

$$F(L) = \left(\sum_{k=1}^{L} \left(\sigma^2_{\mathrm{PC},k} \right) \right) \cdot \left(\sum_{k=1}^{K} \left(\sigma^2_{\mathrm{PC},k} \right) \right)^{-1} . \tag{40}$$

Fig. 25 (b) depicts this metric for our sample data variance over the leading 25 dimensions. One can see that the leading five PC bands capture over 99 % and the leading 10 bands over 99.5 % of the total variance. The answer to what fraction is high "enough" is ambiguous and depends on the application.

As a final remark, it is important to note that subspace transformations are not generically applicable to all hyperspectral image analysis problems and applications, but should be used with caution. In particular, hyperspectral data analysis via PCA can fail if the materials of interest are sparsely distributed in the data or the spectral features that constitutes those materials are hardly distinguishable from other materials' spectral features in the same wavelength range.

2.1.7 Satellite Missions

Satellite optical remote sensing has become an increasingly essential element of everyone's daily life. Mobile software such as Google Maps and Apple Maps use commercial multispectral imagery acquired by the satellite fleet owned by DigitalGlobe Inc., among other data. Today, this constellation includes the WorldView-1 to 4 and GeoEye-1 satellites.

Supplemented by Tab. 2, Tab. 1 lists and summarizes specifications of selected current, recent and planned civilian hyperspectral and multispectral satellite missions. Particular emphasis is given to the spatial and spectral sensor characteristics relevant to multisensor multiresolution data fusion. This table confirms and quantifies the aforementioned trade-off between spatial and spectral resolution. While all listed hyperspectral instruments have a GSD larger than or equal to 30 m, many multispectral and panchromatic imagers reach GSDs of less than or equal to 2 m and 1 m, respectively. Moreover, Tab. 1 reveals that there are more hyperspectral missions planned than there exist today. The only civilian hyperspectral mission covering both VNIR and SWIR, that has been operational in space, was the Earth Observing-1 (EO-1) satellite, which was launched in 2000 and terminated in 2017. After HYPERION onboard EO-1 has stopped delivering data, there is now a period during which no satellite-based hyperspectral SWIR data is available until the first of PRISMA, HISUI or ENMAP will be launched. This fact indicates that there is a significantly larger market and stronger commercial interest for high-spatial resolution multispectral imagery than there is for hyperspectral data. As one achievement of this work is the spatial resolution enhancement of hyperspectral imagery, it implicitly aims at opening new fields of applications and, hence, increasing the public and commercial interest in hyperspectral satellite missions.

2.2 Multiresolution Data Fusion

Due to physical, technological and financial limitations and inevitable trade-offs in Earth observation, each satellite mission is uniquely designed to meet specific objectives. Hence, different satellite-based remote sensing systems capture different information of the Earth's surface. Similarly, it is common to mount multiple complementary sensors onboard remote sensing satellites (see Tab. 1). Multi-sensor data can facilitate a more complete understanding of an area on the ground. The process of combining information from different sensors is referred to as *data fusion*. If those sensors are imaging instruments, one speaks of (remote sensing) *image fusion*. Moreover, image fusion can be categorized into image/pixel-level, feature-level, and decision-level fusion.

The fusion algorithms presented in this thesis belong to the lowest-, i.e. pixel-level image fusion category, and the corresponding multisensor data features complementary spectral and spatial characteristics. More precisely, there are two specific image fusion problems this

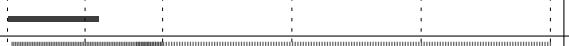
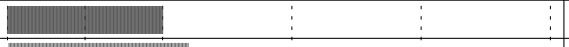
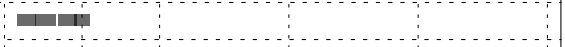
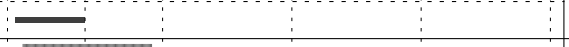
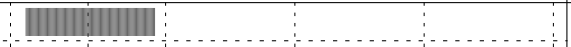
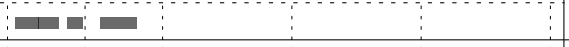
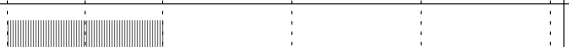
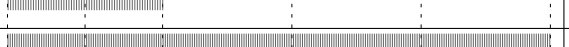
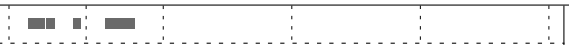
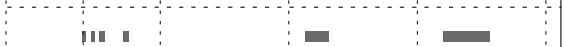
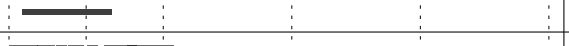
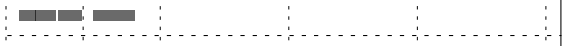
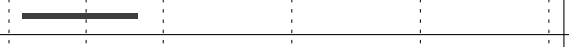
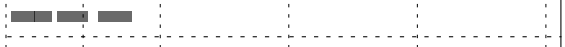
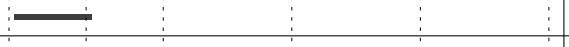
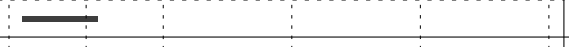
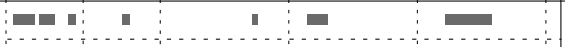
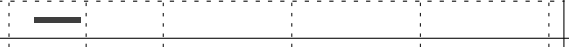
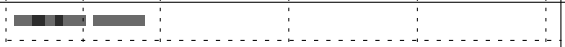
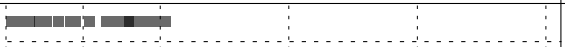
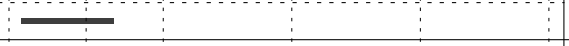
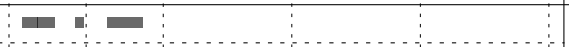
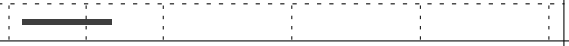
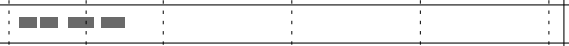
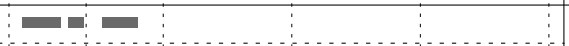
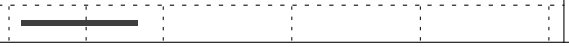
Mission (sensor) name	**Year of launch**	**Sensor class** (band subset)	**Swath width** [km]	**Numb. of bands**	**GSD** [m]	**Spectral coverage** [nm] 400 700 1000 1500 2000 2500	**Ref.** (link to Tab. 2)
Hyperspectral missions							
HISUI	2019	HS	20	185	30		[1-3]
PRISMA	2018	HS	30	267	30		[4-6]
		Pan		1	5		
EnMAP	2018	HS	30	244	30		[7]
DESIS	2017	HS	≈30	235	30		[8,9]
Resurs-P No.2	2014	HS	30	130	30		[10,11]
		MS	38	5	3-4		
		Pan		1	1		
HJ-1A	2008	HS	50	115	100		[12]
		MS	360	4	30		
PROBA (CHRIS)	2001	HS	14	62	34		[13]
EO-1 (HYPERION + ALI)	2000	HS	7.7	242	30		[14]
		MS	37	9	30		
		Pan		1	10		
Multispectral missions							
Sentinel-2A / Sentinel-2B	2015 / 2017	MS (2-4,8)	290	4	10		[15-17]
		MS(5-7,8a, 11,12)		6	20		
		MS(1,9,10)		3	60		
WorldView-4	2016	MS	13.1	4	1.24		[18]
		Pan		1	0.31		
WorldView-3	2014	MS(VNIR)	13.1	8	1.24		[19]
		MS(SWIR)		8	3.7		
		MS(CAVIS)		12	30		
		Pan		1	0.31		
SkySat-2	2014	MS	8	4	2		[20]
		Pan		1	0.9		
Deimos-2	2014	MS	12	4	4		[21,22]
		Pan		1	1		
Spot-6 / Spot-7	2012 / 2014	MS	60	4	6		[23,24]
		Pan		1	1.5		
Landsat-8 (OLI)	2013	MS	185	8	30		[25]
		Pan		1	15		
Pléiades-1A / Pléiades-1B	2011 / 2012	MS	20	4	2		[26,27]
		Pan		1	0.5		
WorldView-2	2009	MS	16.4	8	1.84		[28,29]
		Pan		1	0.46		
GeoEye-1	2008	MS	15.2	4	1.65		[30,31]
		Pan		1	0.41		
RapidEye	2008	MS	77	5	6.5		[32-34]
QuickBird	2001	MS	8	4	2.4		[35,36]
		Pan		1	0.6		

Table 1. List and specifications of selected relevant optical satellite remote sensing missions and instruments. Particular emphasize is given to the spatial and spectral sensor characteristics representing the key system parameters for the multisensor data fusion methodology discussed in this thesis. The numerical means of referencing in the rightmost column is linked to Tab. 2.

[1]=[Tanii et al., 2016]	[13]=[Chan et al., 2011]	[25]=[Pahlevan et al., 2014]
[2]=[Obata et al., 2016]	[14]=[Middleton et al., 2013]	[26]=[Tinel et al., 2012]
[3]=[Yamamoto et al., 2015]	[15]=[Baillarin et al., 2012]	[27]=[Qayyum et al., 2015]
[4]=[Candela et al., 2016]	[16]=[Spoto et al., 2012]	[28]=[de la Fuente et al., 2013]
[5]=[Bonis et al., 2015]	[17]=[Wiedermann et al., 2014]	[29]=[Aguilar et al., 2014]
[6]=[Stefano et al., 2013]	[18]=[DigitalGlobe, 2016]	[30]=[Aguilar et al., 2014]
[7]=[Guanter et al., 2015]	[19]=[Maglione, 2016]	[31]=[Lein, 2012, p. 77]
[8]=[Eckardt et al., 2015]	[20]=[d'Angelo et al., 2014]	[32]=[Stoll et al., 2012]
[9]=[Müller et al., 2016]	[21]=[Pirondini and Al Marri, 2014]	[33]=[Li et al., 2012]
[10]=[Zelentsov et al., 2016]	[22]=[Pirondini et al., 2014]	[34]=[Zeng et al., 2011]
[11]=[Peshkun, Oct. 20-22, 2014]	[23]=[Alganci et al., 2013]	[35]=[de la Fuente et al., 2013]
[12]=[Zhao et al., 2010]	[24]=[Celik et al., 2015]	[36]=[Lein, 2012, p. 76]

Table 2. References used in Tab. 1.

thesis presents solutions for, namely *pan-sharpening* and *hyperspectral-multispectral image fusion*, which are respectively introduced in Sections 2.2.1 and 2.2.2 below.

Prior to the actual fusion process, the data acquired by different sensors should be individually – or, if possible, jointly – pre-processed by means of radiometric and geometric correction including geocoding [Pohl, 2016] (see Section 2.1.3). The minimum requirement for a successful fusion implementation, however, is that the data sets are co-registered. Although image co-registration could be considered as part of the fusion procedure, it is typically treated separately. That is, image fusion algorithms usually require co-registered input data.

Fig. 26 illustrates how spectrally and spatially complementary fusion input data is acquired by sensors mounted on (a) the same satellite platform and (b) different platforms. It reveals that multi-platform multisensor data fusion requires additional planning and processing, and brings limitations in terms of acquisition dates, location and swath overlap. Moreover, if there is a significant time gap between the dates of multi-platform acquisitions, the fusion results can be negatively affected. Factors such as cloud coverage, seasonal changes and other means of changes on the ground or in the atmosphere may introduce artifacts or other artificial phenomena to the image fusion product.

The fusion product can be considered an image as would be acquired by a synthetic sensor ideally featuring the best of the spectral and spatial properties of both sensors the input data was acquired by. Fig. 27 shows a flowchart of the more general multi-platform multisensor image fusion framework. It suggests that remote sensing image fusion incorporates more system information than the mere input images. Depending on the individual pre-processing levels of the input data, the fusion product is radiometrically corrected, geocoded, noise-reduced data with the higher of both spatial and spectral resolutions.

As will be detailed below, the dimensionality of the fusion product is larger than the cumulative dimensionalities of both input images, which makes the fusion problem underdetermined. Since under-determined problems have infinitely many solutions, quality assessment of the fusion results is a non-trivial issue in remote sensing image fusion for which there exists no universal standard to date. The most common means of evaluation requires the availability of reference data, sometimes loosely called *ground truth*, which the fusion results can be compared to using e.g. similarity metrics. Therefore, semi-simulated fusion input data from high-spatial high-spectral resolution data, as for instance acquired by airborne systems, can greatly facilitate comprehensive and comparative algorithm assessment.

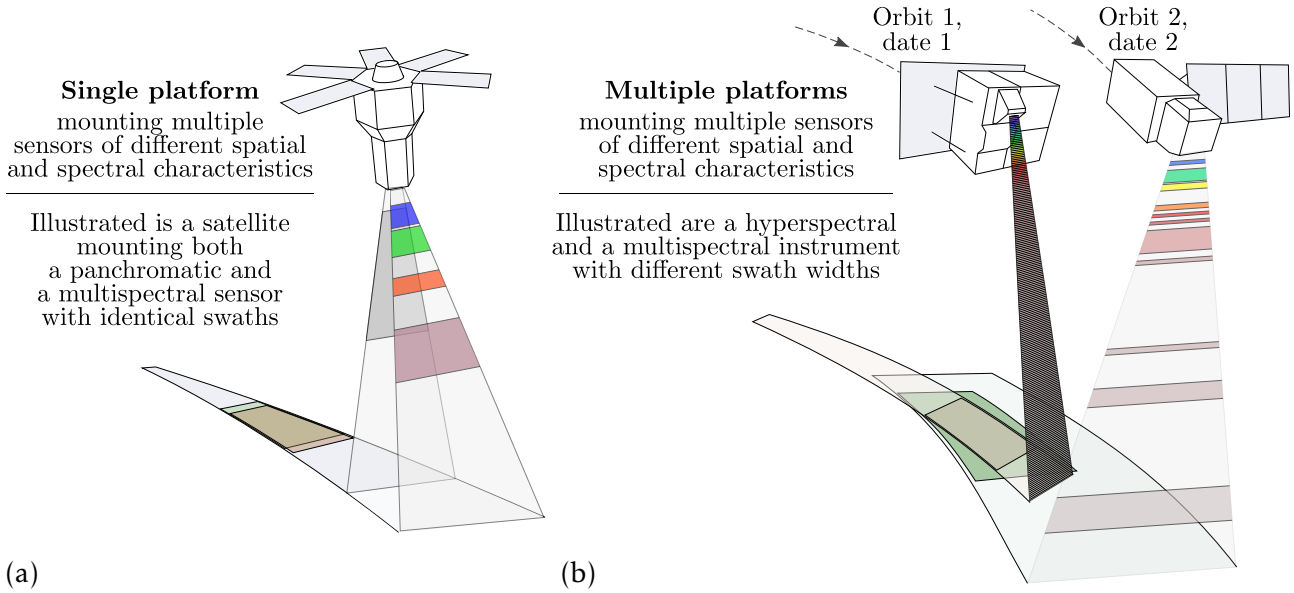


Fig. 26. Multiresolution data acquisition by optical sensors mounted on (a) the same platform and (b) multiple platforms. If mounted on the same platform, complementary instruments are usually designed and mounted in a way that their swaths largely or fully overlap (see the column titled "swath width" in Tab. 1). This facilitates multisensor information extraction, image registration and data fusion. If mounted on multiple platforms – possibly flying on different orbits – data fusion requires precise acquisition planning and makes image fusion preparations including co-registration more difficult. *(Satellite graphics are fictitious, simplified, self-created models inspired by images of the WorldView-4 (left) [Bullock, 2016], mirrored EnMAP (center) [Locherer, 2014] and Sentinel-2 (right) [Drusch et al., 2012] satellites)*

Given the importance of the data simulation and fusion performance assessment, both subjects are separately introduced in Sections 2.2.3 and 2.2.4 below. Finally, the usefulness of multiresolution data fusion is extended on in Sections 2.2.5 and 2.2.6, which respectively discuss potential sensor combinations and applications.

2.2.1 Pan-Sharpening

Pan-sharpening has been the most prominent case of multisensor multiresolution remote sensing data fusion in the past decades, in which the higher-spatial resolution sensor acquires panchromatic – i.e. single-broadband – imagery, and the lower-spatial resolution sensor delivers multispectral data. Brought into context of Fig. 27, the panchromatic sensor's SRF is a single broadband sensitivity function in the visible or VNIR wavelength range, while the multispectral SRFs indicate sensitivity in multiple narrower wavelength regions typically within the VNIR spectral portion (see Tab. 1). Two of many examples of single-platform multispectral-panchromatic sensor combinations are the WorldView-2 and the QuickBird satellites, whose SRFs are depicted in Fig. 13.

One of the factors having a high impact on pan-sharpening performances, is the GSD-ratio, i.e. the GSD of the lower-spatial resolution input sensor relative of the GSD of the higher-spatial resolution input sensor. Larger GSD-ratios typically lead to fusion quality reduction [Zhu, Grohnfeldt, and Bamler, 2016; Yokoya, Grohnfeldt, and Chanussot, 2017]. As can be seen in Tab. 1, the GSDs of sensors onboard the same platform are usually chosen such that the GSD of the multispectral sensor is an even multiple of the panchromatic sensor's GSD. Such setup greatly facilitates the joint geometric correction and image co-registration processes and, thus, the image fusion quality. In most cases, the GSD-ratio is equal to four (see Tab. 1).

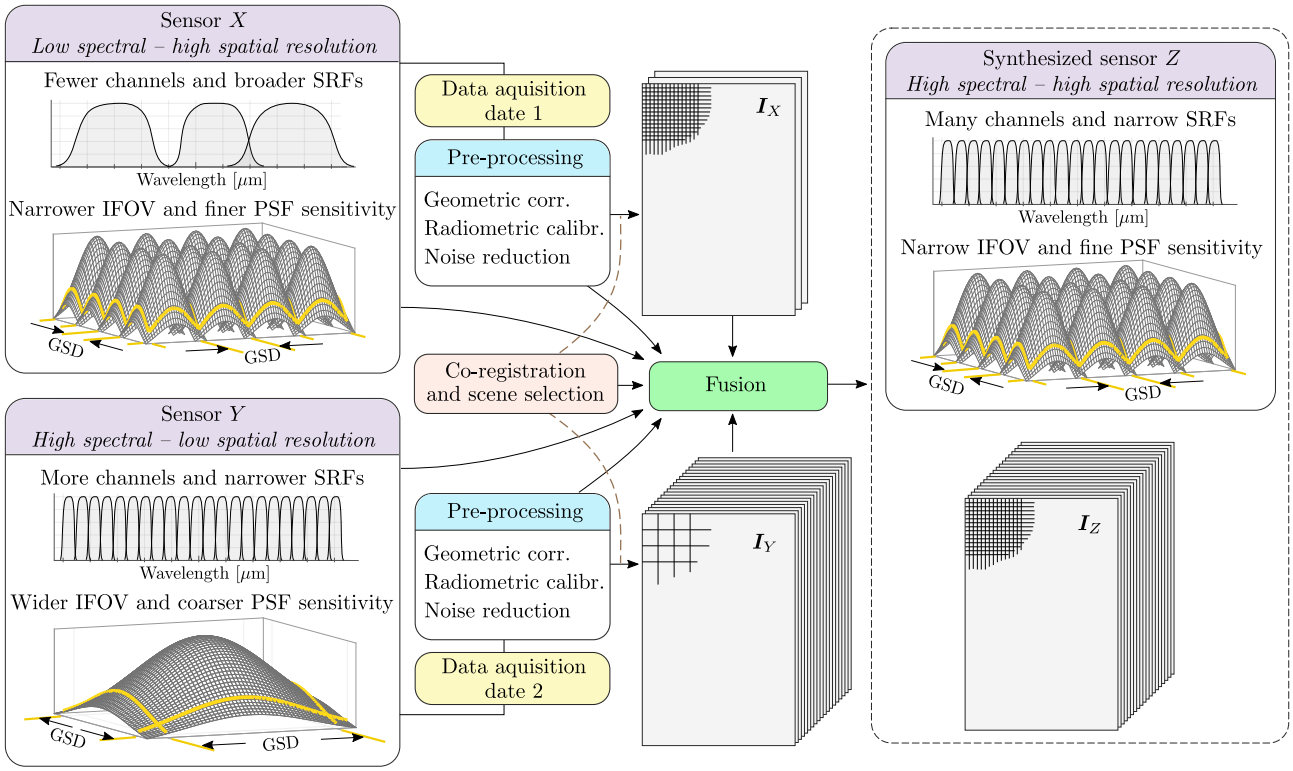


Fig. 27. Flowchart of a general multi-platform multiresolution data fusion framework. In the single-platform multisensor case, the both sensors can acquire the data simultaneously over the same area. Then, the pre-processing and co-registration steps can be conducted jointly for both images. In both cases, the fusion product can be considered an image as would be acquired by an artificial sensor ideally featuring the best of the spectral and spatial properties of both sensors the input data was acquired by. In remote sensing image fusion, system- and pre-processing information are available in addition to the mere input images, which can help improving the fusion product.

The popularity of the pan-sharpening problem has resulted in various categories of pan-sharpening methods as will be extended on in Section 3.1.

2.2.2 Hyperspectral-Multispectral Image Fusion

While the vast majority of current and planned multispectral satellite systems have an additional panchromatic sensor onboard intended for pan-sharpening, systematic *hyperspectral* resolution enhancement using an auxiliary higher-resolution multispectral sensor has been less well established to date. Despite the fact that hyperspectral sensors generally have a coarser spatial resolution than multispectral instruments, and would thus benefit even more from resolution enhancement, there are still operational and planned hyperspectral satellite missions which have only the imaging spectrometer itself on board, such as the PROBA [Chan et al., 2011], DESIS [Müller et al., 2016], HISUI [Tanii et al., 2016] and EnMAP [Guanter et al., 2015] missions. This fact already indicates that there is more to hyperspectral resolution enhancement than a simple adaption of the pan-sharpening framework.

The spectral responses of multispectral imagers, that are operated along with a higher-resolution panchromatic sensor, usually overlap with the broad spectral response of the panchromatic sensor to a large extend. Hyperspectral imagers, on the other hand, may have dozens to hundreds of channels which are not covered by the SRFs of the higher-spatial resolution multispectral sensor. Hence, there may be much information, by means of broad wavelength ranges, collected by the hyperspectral imager, which is not even indirectly captured by the higher-resolution sensor. Moreover, in addition to challenges brought by larger

GSD-ratios between the input sensors, the generally larger difference in spectral information quantity makes the hyperspectral-multispectral data fusion problem more substantial than the pan-sharpening problem.

Another important difference between the two fusion problems is the number and strength of requirements on the fusion products. Many major multispectral mapping applications, such as Google Maps, primarily use three of possibly more multispectral channels to create true-color composites. Such imagery is primarily produced to be easily interpretable by the human eye. In such cases, pan-sharpened imagery is required to be "sharper-looking" color imagery. Slight spectral distortions, as are often introduced by pan-sharpening techniques, are merely or not-at-all noticeable by the human eye. Similarly, for multispectral applications and data analysis purposes that do utilize more or all of the available multispectral channels, including NIR ones, slight spectral distortions may not effect the results, conclusions and decisions strongly. This is partly because in multispectral data analysis and information extraction, spatial properties such as object shapes as well as geometrical structures and textures, are generally accounted for in addition to the spectral properties.

Satellite-based hyperspectral data analysis, methodologies and applications, on the other hand, focus more or solely on spectral characteristics given the poor spatial and high spectral resolutions and the overall mission objectives. Considering the fact the hyperspectral remote sensing aims at identifying and discriminating potentially very similar materials by their spectral profiles, even slight spectral distortions can lead to miss-associations and have more severe negative impacts on the results.

Fig. 28 depicts the concept and benefits of hyperspectral-multispectral data fusion. In contrast to pan-sharpening, which mainly improves the *quality* of the data and products rather than the types of applications, hyperspectral resolution enhancement enables new applications on the global scale which, to date, are possible only locally using higher-spatial resolution (e.g. airborne) hyperspectral data, as will be extended on in Section 2.2.6 below.

2.2.3 Data Simulation

In order to quantitatively assess and compare the quality of images produced by different fusion methods, experiments are typically conducted using simulated semi-real data sets. As will be detailed in Section 2.2.4 below, one popular means of evaluation requires the availability of reference data which the fusion product should – for high scores – be similar to with respect to pre-defined similarity measures. Before introducing the generally applicable simulation framework, which is used in the majority of experiments presented in Sections 4 and 5, it should be noted that data simulation can be arbitrarily comprehensive. The more information about the sensors and desired acquisition condition is available, the more detailed and realistic the data can be modeled. Segl et al., for instance, conducted extensive research on the development of realistic EnMAP and Sentinel-2 data simulation tools as published in [Segl et al., 2012] and [Segl et al., 2015], respectively. Those advanced simulators account for geometric and radiometric distortions as well as sensor specific system noise including striping noise in the case of EnMAP. For detailed descriptions of both the EnMAP and Sentinel-2 data simulators, the interested reader is encouraged to refer to the two aforementioned publications. One of the key experiments in Section 5 is based on simulated EnMAP and Sentinel-2 data provided by Segl et al.

The remainder of this subsection introduces a more generally applicable simulation framework which focusses on the modelling of data sets based on the most essential spectral and spatial system characteristics by means of SRFs, PSFs and statistical sensor noise. This

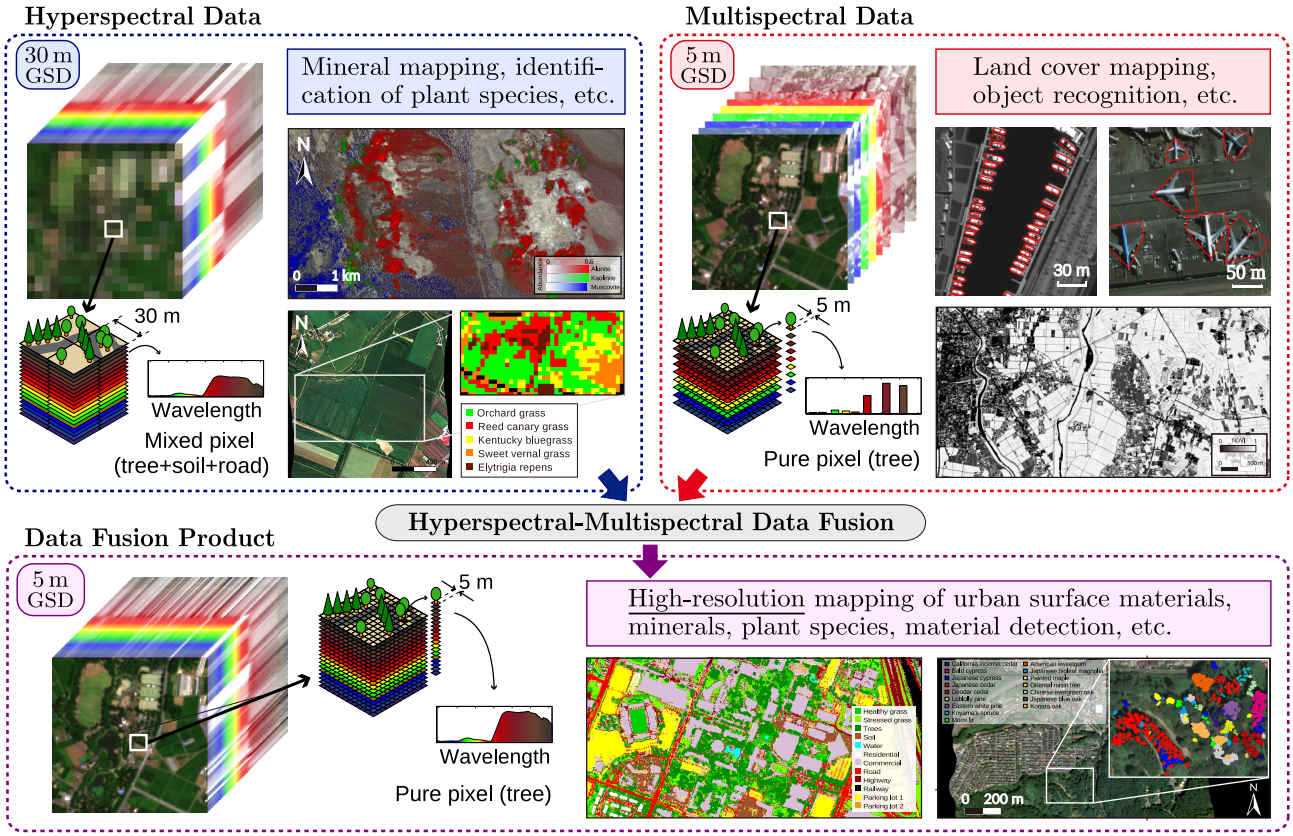


Fig. 28. Concept and benefits of hyperspectral-multispectral data fusion. Combining the high-spatial resolution information and corresponding data analysis tools developed for multispectral data processing with the high-spectral resolution information and corresponding material identification and discrimination capabilities brought by hyperspectral data opens new fields of application for high-spatial high-spectral resolution data possibly on a global scale. *(Graphics after Yokoya, Grohnfeldt, and Chanussot [2017])*

– or simplified versions of this – simulation concept have been commonly applied in remote sensing multiresolution image fusion algorithm testing recently [Yokoya, Grohnfeldt, and Chanussot, 2017; Zhu, Grohnfeldt, and Bamler, 2016; Loncan et al., 2015; Simões et al., 2015; Grohnfeldt and Zhu, 2015b].

Let $\boldsymbol{I}_A \in \mathbb{R}^{N_b^A \times N_p^A}$ be an image acquired by a sensor indexed by A with N_b^A and N_p^A denoting the image's total numbers of spectral bands and spatial pixels, respectively. Moreover, let $SRF_k^A : \mathbb{R}_+ \to \mathbb{R}_+$ denote the relative SRF of sensor A in band $k \in \{1,\dots,N_b^A\}$ with $\int SRF_k^A = 1$. Similarly, let $PSF_i^A : \mathbb{R}_+^2 \to \mathbb{R}_+$ denote the relative PSF of the detector element corresponding to pixel $i \in \{1,\dots,N_p^A\}$ with $\int PSF_i^A = 1$, whose peak response lies around the geometrical center $(u_i^A, v_i^A) \in \mathbb{R}^2$ of pixel i. With this terminology at hand, spectral and spatial transformation matrices, as introduced in Eq. (34) and (35), can be derived to synthesize data of lower spectral or spatial resolutions.

Let $\boldsymbol{I}_O$ denote the "original" multiband image which the fusion input data set will be simulated from. This data should have, ideally, been pre-processed according to a procession chain similar to the one shown in Fig. 11 on page 22. If $\boldsymbol{I}_X$ and $\boldsymbol{I}_Y$ respectively denote the lower-spectral and lower-spatial resolution images we aim to simulate (compare Fig. 27),

then the following conditions must hold:

$$N_b^X < N_b^Y \leq N_b^O \tag{41}$$
$$N_p^Y < N_p^X \leq N_p^O \tag{42}$$

Moreover, the supports of all SRFs of the simulated sensor X should at least partially intersect with supports of the original sensor's SRFs in order not to generate bands containing only zeros. That is, it is reasonable to assume that

$$\forall k \in \{1,\ldots,N_b^X\} \exists l \in \{1,\ldots,N_b^O\} \text{ s.t. } \operatorname{supp}\left(SRF_k^X\right) \cap \operatorname{supp}\left(SRF_l^O\right) \neq \emptyset. \tag{43}$$

Next, in order to create a lower-spectral resolution image $\boldsymbol{I}_X \in \mathbb{R}^{N_b^X \times N_p^O}$ from $\boldsymbol{I}_O$, we define the spectral transformation matrix $\tilde{\boldsymbol{R}}^{O,X} \in \mathbb{R}^{N_b^X \times N_b^O}$ element-wise as follows:

$$\tilde{R}_{k,l}^{O,X} = \left(SRF_k^X * SRF_l^O\right)\left(\lambda_l^O\right) \quad \forall k \in \{1,\ldots,N_b^X\}, l \in \{1,\ldots,N_b^O\} \tag{44}$$

In Eq. (44), the asterisk symbol denotes the convolution operator and λ_l^O denotes the center wavelength of SRF_l^O at FWHM (compare Fig. 9 on page 15). Similarly, we define a spatial transformation matrix $\tilde{\boldsymbol{S}}^{O,Y}$ which will be used to simultaneously blur and down-sample $\boldsymbol{I}_O$:

$$\tilde{S}_{i,j}^{O,Y} = \left(PSF_i^Y * PSF_j^O\right)\left(u_j^O, v_j^O\right) \quad \forall i \in \{1,\ldots,N_p^Y\}, j \in \{1,\ldots,N_p^O\} \tag{45}$$

The facts that any PSF can be assumed to be responsive in its IFOV and corresponding pixel, and that simulated images should not contain pixels that lie outside of the geometrical area covered by $\boldsymbol{I}_O$, imply the PSF version of the condition stated for the sensors' SRFs in Eq. (43).

Similarly to assuming that the continuous relative spectral and spatial response functions are normalized by their integral, the discrete response matrices are normalized accordingly:

$$R_{k,l}^{O,X} = \tilde{R}_{k,l}^{O,X}\left(\sum\nolimits_{q=1}^{N_b^O} \tilde{R}_{k,q}^{O,X}\right)^{-1} \quad \forall k \in \{1,\ldots,N_b^X\}, l \in \{1,\ldots,N_b^O\} \tag{46}$$
$$S_{i,j}^{O,Y} = \tilde{S}_{i,j}^{O,Y}\left(\sum\nolimits_{p=1}^{N_p^O} \tilde{S}_{p,j}^{O,Y}\right)^{-1} \quad \forall i \in \{1,\ldots,N_p^Y\}, j \in \{1,\ldots,N_p^O\} \tag{47}$$

If sensor Y has a lower spectral resolution than the original sensor O, in which case the second inequality in Eq. (41) becomes a strict "$<$", then we need an additional spectral degradation matrix $\boldsymbol{R}^{O,Y}$ to synthesize imagery of sensor Y. Similarly, if X has a lower spatial resolution than sensor O, an additional spatial degradation matrix $\boldsymbol{S}^{O,X}$ is created similarly to $\boldsymbol{S}^{O,Y}$ in Eqs. (45) and (47).

Moreover, the spatially and spectrally degraded images are corrupted by additive noise terms $\boldsymbol{E}_X$ and $\boldsymbol{E}_Y$, which can be composed of various signal dependent and independent system errors and noise contributions (see Section 2.1.3).

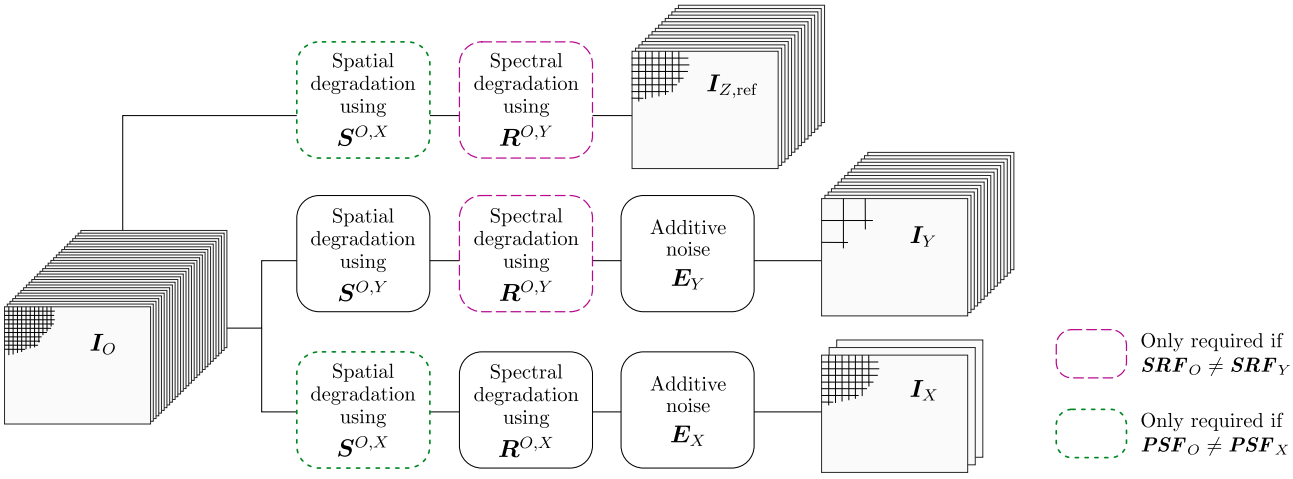


Fig. 29. Flowchart of the data simulation concept used to simulate lower-spatial and lower-spectral resolution fusion input data, as well as reference data from high-spatial high-spectral resolution preprocessed original remote sensing imagery.

Finally, the two input images and the reference data are simulated as follows:

$$\boldsymbol{I}_X = \boldsymbol{R}^{O,X}\boldsymbol{I}_O\boldsymbol{S}^{O,X} + \boldsymbol{E}_X \tag{48}$$

$$\boldsymbol{I}_Y = \boldsymbol{R}^{O,Y}\boldsymbol{I}_O\boldsymbol{S}^{O,Y} + \boldsymbol{E}_Y \tag{49}$$

$$\boldsymbol{I}_{Z,\text{ref}} = \boldsymbol{R}^{O,Y}\boldsymbol{I}_O\boldsymbol{S}^{O,X} \tag{50}$$

If sensors Y and O feature the same spectral characteristics in terms of number of bands and SRFs, then $\boldsymbol{R}^{O,Y}$ is set to the identity matrix $\mathbb{1}^{N_b^Y \times N_b^Y}$. Similarly, if sensors X and O feature the same spatial characteristics in terms of number of pixels and PSFs, then $\boldsymbol{S}^{O,X}$ is set to $\mathbb{1}^{N_p^X \times N_p^X}$.

Fig. 29 depicts the flowchart of the above-described simulation procedure. The symbols $\boldsymbol{SRF}_A$ and $\boldsymbol{PSF}_A$, $A \in \{O, X, Y\}$, in the lower right corner refer to the full sets of functions $\{SRF_1^A, \ldots, SRF_{N_b^A}^A\}$ and $\{PSF_1^A, \ldots, PSF_{N_p^A}^A\}$, respectively.

Finally, note that $\boldsymbol{R}^{O,A}$ *equally* transforms all spectral profiles each of which corresponds to one spatial pixel in the image and $\boldsymbol{S}^{O,A}$ *equally* transforms all images each of which corresponds to one spectral band. This modelling only approximates the real case in which the SRFs of all detector elements corresponding to the same band and the PSFs of all detector elements corresponding to the same pixel but different bands may differ from each other. There are also instruments which comprise bands of significantly different spatial resolutions in terms of GSD, such as Sentinel-2 or WorldView-3. In those cases, all bands that have the same GSD are simulated jointly as one sensor following the procedure described above.

2.2.4 Fusion Quality Assessment

The quality assessment of image fusion products has been an actively debated research topic for over a decade, for which there are still no universally accepted standards yet [Alparone et al., 2004, 2008; Khan et al., 2009; Li et al., 2010; Liu et al., 2011; Yuhendra et al., 2012; Makarau et al., 2012; Pistonesi et al., 2015]. For real data fusion products there is generally no "truth" or reference data available which the product could be compared to. One could say that, if the best-to-be-expected product would be known, image fusion would not be needed. In order yet to be able to assess and compare the performance of different fusion

methods, it is common to fuse semi-real data sets that have been derived from real remotely sensed data following a simulation procedure equal or similar to the one described in the previous section. This approach has the advantage of reference data being available which the fusion product should be close to, with respect to some similarity metrics, for a positive quality assessment. Some of the widely accepted assessment measures, all of which will be used in Sections 4 and 5, are introduced in the following [Du et al., 2007; Vivone et al., 2015; Yokoya, Grohnfeldt, and Chanussot, 2017].

Let $\boldsymbol{Z} = [\boldsymbol{z}_1, \ldots, \boldsymbol{z}_{N_b}]^T = [\mathbf{z}_1, \ldots, \mathbf{z}_{N_p}] \in \mathbb{R}^{N_b \times N_p}$ denote a reference image composed of N_b spectral bands and N_p spatial pixels. Moreover, let $\hat{\boldsymbol{Z}}$ denote an equally sized fusion product [6]. The following evaluation metrics measure different aspects of similarity between $\boldsymbol{Z}$ and $\hat{\boldsymbol{Z}}$.

Peak signal-to-noise ratio (PSNR): In image processing, the PSNR describes the ratio between the maximum measured (or possible) intensity value and the magnitude of the residual error [Yokoya, Grohnfeldt, and Chanussot, 2017]. In order to avoid a strong dependency on the dynamic range of the data, the PSNR is often expressed logarithmically in terms of decibel:

$$PSNR(\boldsymbol{z}_k, \hat{\boldsymbol{z}}_k) = 20 \cdot \log_{10}\left(\frac{\max(\boldsymbol{z}_k)}{RMSE(\boldsymbol{z}_k, \hat{\boldsymbol{z}}_k)}\right) \text{[dB]} \tag{51}$$

In Eq. (51), $\max(\boldsymbol{z}_k)$ is the maximum element in $\boldsymbol{z}_k$ and $RMSE(\boldsymbol{z}_k, \hat{\boldsymbol{z}}_k)$ denotes the *root mean square error* between the two vectors $\boldsymbol{z}_k$ and $\hat{\boldsymbol{z}}_k$. The latter quantity is defined as

$$RMSE(\boldsymbol{z}_k, \hat{\boldsymbol{z}}_k) = \left(\frac{\|\boldsymbol{z}_k - \hat{\boldsymbol{z}}_k\|_2^2}{N_p}\right)^{1/2}. \tag{52}$$

Larger PSNR values indicate higher spatial image reconstruction fidelity.

Erreur relative globale adimensionnelle de synthèse (ERGAS): The ERGAS is a literal measure for the relative dimensionless global error in synthesis. In contrast to PSNR, it is explicitly defined for *multiband* image fusion assessment as [Wald, 2000]

$$ERGAS(\boldsymbol{Z}, \hat{\boldsymbol{Z}}) = \frac{100}{F_{ds}}\left(\frac{1}{N_b}\sum_{k=1}^{N_b}\left(\frac{RMSE(\boldsymbol{z}_k, \hat{\boldsymbol{z}}_k)}{\overline{\boldsymbol{z}_k}}\right)^2\right)^{1/2}, \tag{53}$$

where F_{ds} denotes the GSD ratio between the lower- and higher-spatial resolution fusion input image and $\overline{\boldsymbol{z}_k}$ is the mean value of the vector $\boldsymbol{z}_k$. The lower the ERGAS value, the better the fusion product. Note that, if applied to individual bands $\boldsymbol{z}_k$ and $\hat{\boldsymbol{z}}_k$ instead of $\boldsymbol{Z}$ and $\hat{\boldsymbol{Z}}$, ERGAS can be used to assess the spatial quality of the fusion result in band k, similarly to PSNR [Zhu, Grohnfeldt, and Bamler, 2016].

Correlation coefficient (CC): The CC quantifies the correlation between two single-band image vectors $\boldsymbol{z}_k$ and $\hat{\boldsymbol{z}}_k$. Its definition is stated in Eq. (32), but repeated here for convenience:

$$CC(\boldsymbol{z}_k, \hat{\boldsymbol{z}}_k) = \frac{\sigma_{z_k, \hat{z}_k}}{\sigma_{z_k}\sigma_{\hat{z}_k}} \tag{54}$$

The CC takes values between -1 and 1, where higher correlation coefficients indicate better reconstruction results.

$Q2^n$: The $Q2^n$ is a generalization of the *universal quality index* (UIQI) [Wang and Bovik,

[6] The full reference image and the fusion product are usually denoted by $\boldsymbol{I}_{Z,\text{ref}}$ and $\boldsymbol{I}_Z$ in this thesis. The letters $\boldsymbol{Z}$ and $\hat{\boldsymbol{Z}}$ are used in this section only temporarily to facilitate readability.

2002], which is defined for monochromatic images as

$$UIQI(\boldsymbol{z}_k, \hat{\boldsymbol{z}}_k) = \frac{\sigma_{z_k,\hat{z}_k}}{\sigma_{z_k}\sigma_{\hat{z}_k}} \cdot \frac{2\overline{z_k}\,\overline{\hat{z}_k}}{(\overline{z_k})^2 + (\overline{\hat{z}_k})^2} \cdot \frac{2\sigma_{z_k}\sigma_{\hat{z}_k}}{\sigma_{z_k}^2 + \sigma_{\hat{z}_k}^2}, \tag{55}$$

and was proposed as an extension of the $Q4$ metric [Alparone et al., 2004] to multiband images of more than four bands. By treating spectral profiles as hypercomplex numbers it defines a metric which jointly measures spectral and spatial distortions. Just like UIQI, $Q2^n$ maps onto the unit interval with 1 indicating the best and 0 indicating the worst possible fusion product quality. A detailed description of the $Q2^n$ index can be found in [Garzelli and Nencini, 2009].

Spectral angle mapper (SAM): The SAM indicates the relative spectral distortion at each spatial pixel by measuring the angle between the corresponding spectral profiles $\mathbf{z}_i$ and $\hat{\mathbf{z}}_i$, $i = 1,\ldots,N_p$, as follows [Kruse et al., 1993]:

$$SAM(\mathbf{z}_i, \hat{\mathbf{z}}_i) = \arccos\left(\frac{\mathbf{z}_i^T \hat{\mathbf{z}}_i}{\|\mathbf{z}_i\|_2 \|\hat{\mathbf{z}}_i\|_2}\right) \tag{56}$$

The narrower the angle, the closer is the fusion result to the reference, disregarding a possible amplitude factor. While the average SAM over all pixels, i.e.

$$SAM_{\text{avg}}(\boldsymbol{Z}, \hat{\boldsymbol{Z}}) = \sum_{i=1}^{N_p} SAM(\mathbf{z}_i, \hat{\mathbf{z}}_i), \tag{57}$$

gives an indication of the overall spectral distortion, the pixel-wise SAM information can be depicted as a monochromatic image in order to inspect the spectral distortion distribution across the image.

Those of the above-described metrics that are applicable to individual band pairs $(\boldsymbol{z}_k, \hat{\boldsymbol{z}}_k)$, can be used to assess the band- and wavelength-dependent fusion performance.

Aside from the assessment results based upon these objective statistical similarity measures, the fusion products can be variably suitable for different applications. Since, in many applications, optical remote sensing data is still used to create maps which are, at some point, visually inspected by a person, a qualitative and, to some extend, subjective assessment of the fusion product remains a valuable means of quality assurance. In addition to the SAM "image", also the pixel-wise L_2-norm of the difference image $(\boldsymbol{Z} - \hat{\boldsymbol{Z}})$ can be used to visually analyze the error distribution in the image and possibly identify artifacts [Yokoya, Grohnfeldt, and Chanussot, 2017].

As mentioned in the previous sections, classification maps are standard products, especially in hyperspectral remote sensing. The accuracy of such maps and map-creating classifiers can be quantitatively calculated if ground truth information is available. If the same classification algorithm is used to classify images produced by different fusion algorithms, the comparative classification accuracies add another complementary means of fusion quality evaluation [Chan et al., 2011; Yokoya, Grohnfeldt, and Chanussot, 2017]. Such experiments are conducted in Section 5 to assess the accuracy improvement of classification maps after performing hyperspectral resolution enhancement.

As a final remark, it should be noted that a suitable selection of fusion quality assessment measures should be made according to the application the fusion products are intended to

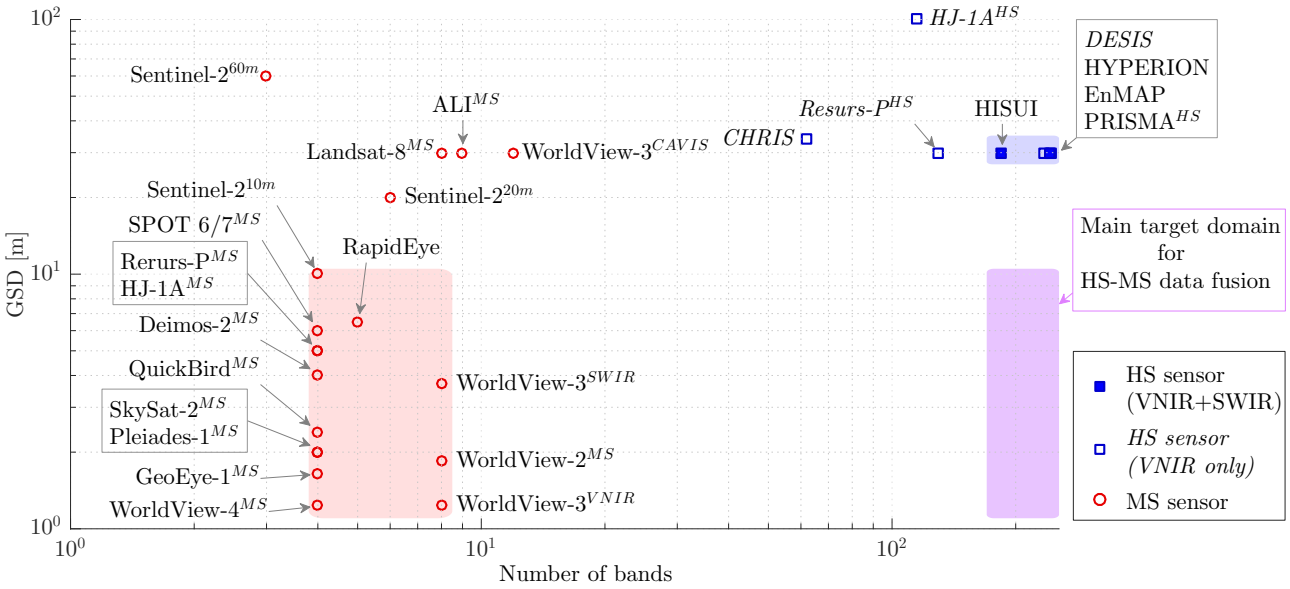


Fig. 30. Illustration of the trade-off between the spectral and spatial resolution of selected recent, current and future satellite-based electro-optical multi- and hyperspectral remote sensing instruments brought by physical and technological hardware limitations. In this figure, the spectral and spatial resolutions are loosely quantified by the number of bands and the ground sampling distance (GSD), respectively. Furthermore, the graphic indicates the target spectral-spatial characteristics of what can be thought of as synthetic sensors generated by multiresolution data fusion.

be used for. For instance, SAM- and classification-based quality metrics are most valuable if the fusion products are intended to be used for material identification purposes. If the products are pan-sharpened RGB imagery primarily intended to be used for visual analyses, then quality assessment based on visual inspection is clearly indispensable.

2.2.5 Missions and Potential Sensor Combinations

As can be seen from Tab. 1, some of the recent, current and future civilian remote sensing satellites carry multiple complementary electro-optical sensors while others are equipped with only one multi- or hyperspectral instrument. Except for Sentinel-2 and RapidEye, all listed multispectral sensors are supplemented by a broadband panchromatic sensor, whose primary purpose is to provide complementary higher-spatial resolution data that can be fused with the multispectral imagery for resolution enhancement. Similarly, some of the listed hyperspectral missions carry supplementary sensors of higher spatial and lower spectral resolution. Clearly, simultaneous acquisition of data by sensors mounted onboard the same platform from nearly identical viewing angles significantly facilitates data fusion preparations in terms of planning, joint data pre-processing and image co-registration, compared to the case in which fusion input data is acquired by sensors onboard different platforms (see Fig. 26). Hence, all single-platform multisensor satellite missions are natural potential target data sources for multiresolution image fusion. It can be assumed that all modern satellite missions holding both a multispectral and a higher-spatial resolution panchromatic sensor are specifically designed in a way that allows for a relatively straightforward preparation of pan-sharpening input data. This is not necessarily the case for missions carrying both a hyperspectral and a multispectral sensor as will be discussed in the following.

The EO-1 mission, for instance, carries both the 30-m GSD hyperspectral HYPERION and the dual-sensor 30-m multispectral and 10-m panchromatic Advanced Land Imager (ALI) instruments. Both HYPERION and ALI were individually intended as technology demon-

strations for follow-up satellite missions including the Operational Land Imager (OLI) onboard the Landsat Data Continuity Mission's (LDCM) Landsat-8 satellite and the EnMAP hyperspectral satellite mission, while the lifespan of the EO-1 satellite was initially estimated to be one year [Middleton et al., 2013]. Apart from the individual initial objectives of the two instruments, the fact that ALI's multispectral and HYPERION's hyperspectral sensor both have the same GSD of 30 m additionally indicates that the EO-1 mission was *not* designed for pixel-based fusion of HYPERION and ALI data. To the best knowledge of the author, there is very limited literature discussing experiments on full-resolution HYPERION and ALI panchromatic data fusion [Capobianco et al., 2007; Nikolakopoulos, 2009], [Alparone et al., 2015, p. 189]. Similarly, there is no published work on the fusion of real multisensor Resurs-P data and only one paper addressing HJ-1A multisensor data fusion [Xiong et al., 2014] in the international literature to date.

The lack of real-data experiments for systematic operational satellite-based hyperspectral resolution enhancement is a known issue [Yokoya, Grohnfeldt, and Chanussot, 2017]. As will be discussed in Section 3, there are increasing research activities on data fusion algorithm development for hyperspectral resolution enhancement. However, experiments are usually conducted using simulated input data that was derived from high-spatial resolution airborne acquisitions. In order to fully exploit the potentials and possibilities that will be brought by the upcoming hyperspectral satellite missions, more efforts need to be made to develop solutions for the fusion of data acquired by specific single- and multiplatform sensor combinations. The DESIS, EnMAP and HISUI space-based hyperspectral instruments, for instance, will be launched without companion higher-spatial resolution sensors. In contrast, the Italian PRISMA spacecraft will be equipped with both a hyperspectral and a higher-spatial resolution panchromatic sensor. Before the launch of those upcoming hyperspectral programs, mission-specific solutions for resolution enhancement should be developed in order to allow for applications which require, or are facilitated by, a high spatial resolution.

Fig. 30 illustrates the trade-off between the spectral and spatial resolution of selected recent, current and future satellite-based electro-optical multi- and hyperspectral remote sensing instruments, where the spectral and spatial resolutions are loosely quantified by the number of bands and the GSD, respectively. Furthermore, the graphic indicates the target spectral-spatial characteristics of what can be thought of as synthetic sensors generated by multiresolution data fusion. Fusion products featuring a high spatial resolution and hyperspectral information in both the VNIR and SWIR portion of the spectrum are particularly beneficial for many applications, which is why the purple area on the lower-right of Fig. 30 covers only a sub-range of all possible sensor combinations.

Finally, there is very little literature about cross-platform multiresolution optical data fusion to date. To the best knowledge of the author, the only published cross-platform multiresolution data fusion experiments involving real hyperspectral satellite-based data, are based on HYPERION and ASTER multispectral data [Yokoya et al., 2013; Licciardi et al., 2014]. The vast majority of publications on multiresolution data fusion still address the single-platform pan-sharpening problem. According to Pohl [2016], the most popular missions for multiresolution image fusion in the remote sensing literature in the past two decades have been Landsat and IKONOS, followed by Quickbird and SPOT. Even though fusion of hyperspectral data with higher-spatial resolution imagery has been receiving increasing attention from the hyperspectral remote sensing signal processing community, concrete solutions for specific missions, possibly involving cross-platform sensor combinations still offers room for development. Section 5 contains experiments on the fusion of end-to-end simulated EnMAP data [Segl et al., 2012] and 10-m GSD Sentinel-2 data, among others.

2.2.6 Applications

As explained above and indicated in Figs. 27, 28 and 30, the multiresolution fusion products discussed in this work are intended to feature the spatial characteristics of the higher-spatial resolution input image in terms of GSD and spatial responsivity, and the spectral information of the higher-spectral resolution sensor in terms of number of bands and spectral responsivity. Hence, the fusion process can be thought of as spatial resolution enhancement of optical imagery using auxiliary higher-spatial resolution data. With this in mind, applications can be broadly grouped into the following two categories:

New applications: Applications, which would *not be possible* by using the individual input images due to e.g. coarse spatial resolution or lack of spectral information. Physical constraints and the current technological status limit such high-resolution data acquisitions to *lower altitudes*, i.e. airborne or UAV-based remote sensing. One of the main problems of airborne acquisitions, however, is the limited operating flexibility in terms of scheduling and accessibility of areas that are either far away or contain restricted airspace (e.g. many urban areas). Satellite-based remote sensing allows for a *continuous global coverage* which makes it an optimal source for long-term monitoring, prompt disaster assessment and urban mapping. The resolution enhancement capacity of multisensor data fusion combines the best of both airborne and satellite-based remote sensing. Specific examples for applications will be given below.

Facilitation of – and quality improvement for – existing applications: The majority of applications that are principally possible without data fusion would benefit from higher-resolution data. For instance, the accuracy and quality of products, including maps and higher-level information, derived from optical remote sensing data, can be enhanced by multiresolution image fusion.

From Tab. 1 one can observe that the GSD-ratios between current multispectral and panchromatic sensors onboard the same satellite platforms lie between 2 and 4. This ratio determines the spatial resolution enhancement factor for the corresponding multispectral imagery. On the other hand, when looking at Fig. 30, one can see that the GSD-ratios between potential target hyperspectral-multispectral sensor combinations vary between 3 (e.g. combining EnMAP and Senstinel-2) and around 24 (e.g. combining EnMAP and WorldView-4). Considering the presumably larger GSD-ratios in hyperspectral-multispectral fusion on the one hand and the problematic for some hyperspectral applications of having multiple material spectra mixed in single pixels at an original GSD of 30 m on the other hand, it is obvious that a large portion of hyperspectral resolution enhancement applications belong to the first of the two above-described categories, while pan-sharpening is primarily useful for data quality enhancement [Noviello et al., 2013; Yuhendra et al., 2011; Bovolo et al., 2010].

Recently, Pohl [2016] conducted a comprehensive survey on remote sensing image fusion applications based on around 350 relevant peer-reviewed journal publications. Therein, she confirmed that pan-sharpening has been the by far most frequently addressed remote sensing image fusion problem in the past two decades while hyperspectral-multispectral data fusion has been represented only marginally. Furthermore, the fusion applications, most of which correspond to pan-sharpening, and their relative occurrence percentages were found to be as follows: Urban applications (26%), change detection (15%), geology (12%), agriculture (9%), vegetation (7%), forestry (7%), land cover and use (5%), mapping (3%), hazards and disasters (3%), coastal zones (3%) and other (9%).

The applications of hyperspectral-multispectral image fusion partially overlap with those listed above for pan-sharpening. Moreover, any hyperspectral remote sensing application

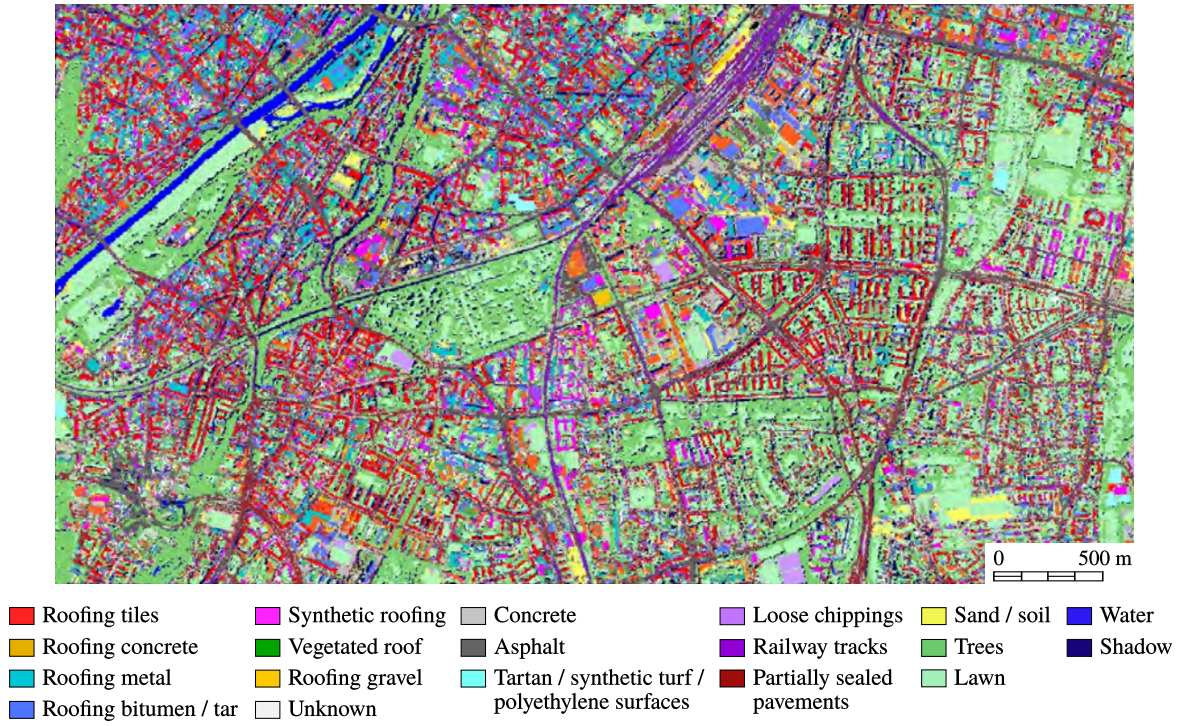


Fig. 31. Classification map of an area in Munich, Germany, based on airborne 4-m GSD high-resolution hyperspectral data acquired by the HyMAP sensor [Cocks et al., 1998] in 2007. The map shows the most dominant of possibly multiple classes in each pixel and was created by Heldens [2010] using a linear unmixing framework. In the scope of this thesis, this high-resolution classification map presents a potential product as can be derived from spatially enhanced satellite-based hyperspectral data. *(All credits for this figure belong to [Heldens, 2010])*

can potentially benefit from spatially enhanced hyperspectral data. A list of general hyperspectral remote sensing applications was provided in Section 2.1.5 above. In the following fields and corresponding publications, the *high spatial resolution* of hyperspectral remote sensing data is specifically utilized: Urban mapping and monitoring [Hardin and Hardin, 2013; Shafri and Taherzadeh, 2012; Huang et al., 2009; Galli and Malinverni, 2003; Marino et al., 2001], urban forestry monitoring in natural disaster zones [Zhang et al., 2011], mapping of post-fire forest regeneration and vegetation recovery [Mitri and Gitas, 2013], detection of plant deceases [Calderón et al., 2013], vegetation management [Frank et al., 2010], mapping of the spatial distribution of vegetation species [Aspinall, 2002], tree species classification [Dalponte et al., 2012], agriculture [López-López et al., 2016; Gonzalez-Dugo et al., 2015; Eddy et al., 2008; Monteiro et al., 2006], analysis of the ecosystem and environment [Legleiter, 2003; Aspinall et al., 2002], archaeology [Boccardo et al., 2004], volcanology [Lombardo et al., 2009], anomaly detection [Goovaerts et al., 2005] and target detection [Nasrabadi, 2014; Yang and Shi, 2016; Bar et al., 2010].

Fig. 31 shows a classification map of an area in Munich, Germany, based on high-resolution hyperspectral data of 4-m GSD [Heldens, 2010]. Even though this data was originally acquired by the airborne HyMAP sensor [Cocks et al., 1998], such a high-resolution classification map can be similarly produced based on spatially enhanced hyperspectral data generated via hyperspectral-multispectral data fusion.

2.3 Sparse Signal Recovery

The analysis and recovery of sparse and compressible signals is a highly active field of research which interconnects a variety of disciplines including applied mathematics, computer science and various engineering domains. Aside from the diversity of hands-on engineering applications, it is particularly interesting for mathematicians as it combines linear algebra, numerical and functional analysis, stochastic and combinatorics. With the seminal works of Candès and Tao [2005]; Candès et al. [2006b,a] and Donoho [2006], who found that sparse signals can be exactly reconstructed at sub-Nyquist sampling rates, the theory of *sparse signal recovery* revived in the scope of *compressed sensing*. Fundamentally, compressed (or *compressive*) sensing involves finding sparse solutions of underdetermined systems. The corresponding theory has received tremendous attention since the above-mentioned works, so that, today, there exist comprehensive literature much of which was nicely summarized by Fornasier [2010]; Eldar and Kutyniok [2012]; Foucart and Rauhut [2013]. Due to the increasing popularity and, hence, visibility of this subject, more and more engineers exploit prior knowledge about sparsity in their mathematical models to gain performance improvements, in particular in terms of reconstruction accuracy. To underline the interdisciplinarity of sparse recovery, I would like to mention, that the research work resulting in this thesis has been conducted in collaboration with the Department for Applied Numerical Analysis of the Technische Universität München. Specifically, Dr. Steffen Peter has provided valuable support in finding and accelerating suitable solving algorithms for the sparse recovery problems modelled in the scope of this work. Having said this, the fundamentals introduced in this section can be formulated in more general terms of infinite dimensional Banach spaces [see e.g. Peter, 2016]. However, given the application-driven context of this dissertation, the following introduction covers only the special case of real-valued variables in finite dimensional spaces.

2.3.1 Notation

For any matrix $\boldsymbol{X} \in \mathbb{R}^{N\times D}$, $N, D \in \mathbb{N}$, let $\boldsymbol{x}_d = [x_{1,d}, \dots, x_{N,d}]^T$ indicate its d-th column and $\mathbf{x}_n = [x_{n,1}, \dots, x_{n,D}]$ its n-th row, respectively. Moreover, let

$$\|\boldsymbol{x}\|_q := \left(\sum_{d=1}^{D} |x_d|^q \right)^{\frac{1}{q}}, \quad 0 < q < \infty, \quad \|\boldsymbol{x}\|_\infty := \max_{d=1,\dots,D} |x_d|, \tag{58}$$

denote the standard l_q-(quasi)-norm in $\mathbb{R}^D$, $D \in \mathbb{N}$. For $0 < q \leq \infty$, let

$$\|\boldsymbol{X}\|_{p,q} := \left(\sum_{n=1}^{N} \|\mathbf{x}_n^T\|_q^p \right)^{\frac{1}{p}}, \quad 0 < p < \infty, \quad \|\boldsymbol{X}\|_{\infty,q} := \max_{n=1,\dots,N} \|\mathbf{x}_n^T\|_q, \tag{59}$$

denote the mixed $l_{p,q}$-(quasi)-norm in $\mathbb{R}^{N\times D}$. In particular, if $q = p = 2$, the norm

$$\|\boldsymbol{X}\|_F := \|\boldsymbol{X}\|_{2,2} \tag{60}$$

denotes the Frobenius-norm in $\mathbb{R}^N$. Moreover, the matrix norm in $\mathbb{R}^{N\times D}$ induced by the l_p-vector norm is denoted by

$$\|\boldsymbol{X}\|_p = \sup\left\{ \boldsymbol{z} \in \mathbb{R}^D \backslash \{\boldsymbol{0}\} \,\middle|\, \|\boldsymbol{X}\boldsymbol{z}\|_p / \|\boldsymbol{z}\|_p \right\}, \tag{61}$$

where, if $p = 2$, $\|\boldsymbol{X}\|_p$ equals the *spectral norm* – i.e. the largest singular value – of $\boldsymbol{X}$.

The support of a vector $\boldsymbol{x} \in \mathbb{R}^N$ is defined as

$$\operatorname{supp}(\boldsymbol{x}) := \left\{ n \in \{1,\dots,N\} \,\middle|\, x_n \neq 0 \right\}. \tag{62}$$

2.3.2 Sparse and Compressible Signals

Let the operator # specify the cardinality of a finite set. Then, a vector $\boldsymbol{x} \in \mathbb{R}^N$ is called s-sparse, if

$$\|\boldsymbol{x}\|_0 := \#\operatorname{supp}(\boldsymbol{x}) \leq s. \tag{63}$$

In Eq. (63), the function $\|\cdot\|_0$ is commonly called "l_0-norm" even though it is not even a quasi-norm [see e.g. Fornasier, 2010]. However, this term is intuitive, because $\|\boldsymbol{x}\|_q^q \to \|\boldsymbol{x}\|_0$ as $q \to 0$ [see e.g. Boche et al., 2015]. The set of all s-sparse vectors in $\mathbb{R}^N$ is denoted by

$$\Sigma_s := \{\boldsymbol{x} \in \mathbb{R}^N \,|\, \|\boldsymbol{x}\|_0 \leq s\}. \tag{64}$$

The number s of non-zero elements in a vector $\boldsymbol{x} \in \mathbb{R}^N$ relative to its dimension N indicates the *level of sparsity*. That is, the smaller the quotient s/N the higher the sparsity level. The question whether or not a vector or signal *is* sparse is not generally answerable, as there is no concrete value for s/N below which a signal is called "sparse". Notwithstanding, if sparse signal recovery tools are utilized, the signals of interest usually have a relatively high sparsity level. Hence, in this work, a signal is called sparse only if it has indeed a high sparsity level, i.e. $s/N \ll 1$.

In engineering applications, signals are often not exactly sparse but *compressible*. That is, most of the vector coefficients are not exactly equal to zero, but "small" in amplitude relative to a few significant entries. More formally, a signal is considered compressible if it has a relatively small *best s-term approximation error* which is defined for $1 \leq p < \infty$ as

$$\varsigma_s(\boldsymbol{x})_p = \min_{\boldsymbol{z}\in\Sigma_s} \|\boldsymbol{x} - \boldsymbol{z}\|_p. \tag{65}$$

2.3.3 Basic Idea of Sparse Recovery

The fundamental idea of sparse recovery is to solve an underdetermined linear system of equations for a sparse solution. That is, given a matrix $\boldsymbol{\Phi} \in \mathbb{R}^{M\times N}$, $M < N$, of rank M and a measurement vector $\boldsymbol{y} \in \mathbb{R}^M$, one is interested in finding – among infinitely many solutions – a sparse vector $\boldsymbol{x}$ which solves the following system:

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x} \tag{66}$$

Depending on the context and application, the matrix $\mathbf{\Phi}$ commonly constitutes a *sensing matrix*, a *dictionary* or the product of both. In this work, $\mathbf{\Phi}$ represents dictionaries. The columns of a dictionary are called *atoms*.

Note that the simple setup described above does not take into account possible system or measurement perturbation. Thus, it is referred to as *noiseless system model*. The noise-aware case will be discussed in Section 2.3.5.

An intuitive approach to finding the sparsest among all possible solutions of Eq. (66) would be to solve the following optimization problem which is referred to as l_0-*norm minimization*:

$$\min_{\boldsymbol{x}\in\mathbb{R}^N} \|\boldsymbol{x}\|_0 \quad \text{s.t.} \quad \mathbf{\Phi}\boldsymbol{x} = \boldsymbol{y} \tag{67}$$

The problem with this approach is that it is known to be *NP-hard*[7] [Mallat and Zhang, 1993; Natarajan, 1995]. In other words, solving Eq. (67) becomes numerically intractable as N and the number s of non-zero coefficients in $\boldsymbol{x}$ get larger. In order yet to be able to efficiently solve Eq. (66) for sparse solutions, Eq. (67) can be relaxed as will be outlined in the following section.

2.3.4 Basis Pursuit and Related Properties

As the l_0 minimization problem stated in Eq. (67) is numerically intractable for larger data, it is desirable to find alternative problem formulations which lead to similarly sparse solutions, but can be solved efficiently for arbitrarily sized systems. Consider substituting the l_0-"norm" in Eq. (67) by the l_p-norm, where $0 \leq p \leq \infty$. Figure 32 illustrates the corresponding solutions for $p = 0$, ½, 1, 2, and ∞ exemplarily in $\mathbb{R}^2$. It can be observed that, for $0 \leq p \leq 1$ the solutions of the illustrated system are sparse. On the other hand, any $p \geq 1$ convexifies the optimization problem. These observations generally hold in $\mathbb{R}^N$ [see e.g. Fornasier, 2010; Elad, 2010; Eldar and Kutyniok, 2012; Foucart and Rauhut, 2013], $N \in \mathbb{N}$, which is why the l_1-norm minimization is a natural candidate for efficiently finding sparse solutions of underdetermined systems. The special case of $p = 1$ is known as *basis pursuit*:

$$\min_{\boldsymbol{x}\in\mathbb{R}^N} \|\boldsymbol{x}\|_1 \quad \text{s.t.} \quad \mathbf{\Phi}\boldsymbol{x} = \boldsymbol{y} \tag{68}$$

In fact, this particular problem has attracted much attention especially among mathematicians and engineers in the past, so that, to date, there exists a comprehensive theory and numerous algorithms for solving Eq. (68) and various similar formulations [see e.g. Eldar and Kutyniok, 2012; Fornasier, 2010; Foucart and Rauhut, 2013]. Some of the fundamental theoretical findings include conditions on $\mathbf{\Phi}$ under which the solutions of the l_0- and the corresponding l_1-norm minimization problem are identical, such as the *Null Space Property* (NSP) and the *Restricted Isometry Property* (RIP). The definitions of both NSP and RIP and related theory can be found e.g. in [Fornasier, 2010, pp. 6-8] and [Eldar and Kutyniok, 2012, p. 17,19]. Put simple, if $\mathbf{\Phi}$ satisfies the NSP, Eqs. (67) and (68) both have the same unique solution. However, since the NSP is hard to prove for the vast majority of matrices, the RIP offers remedy in that it is usually easier to verify at least for some classes of matrices. For instance, random matrices of the following type can be shown to satisfy the RIP (with small RIP constant) with high probability, where the coefficients $\phi_{m,n}$ are independent identically

[7] Roughly speaking, an NP-hard problem is at least as hard to solve as an NP-problem (non-deterministic polynomial time problem), which in turn implicates a computational cost that is polynomial in the dimension of the input.

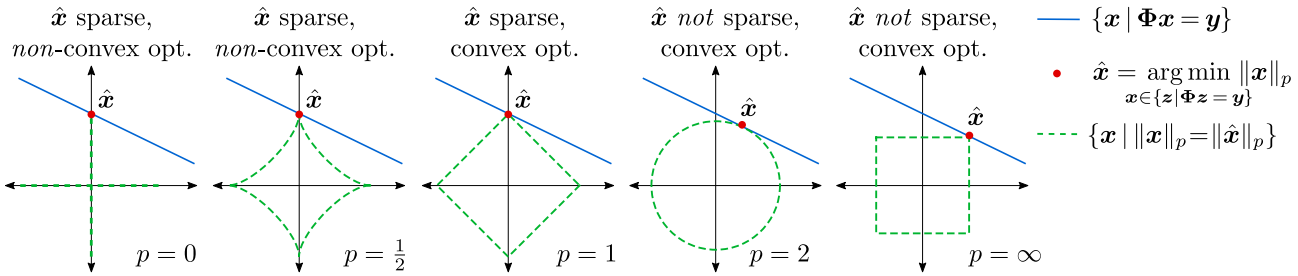


Fig. 32. Illustration of the advantages of l_1-norm minimization; The blue lines indicate the 1-D (linear) solution space of an underdetermined linear system in $\mathbb{R}^2$. The red points indicate those solutions featuring the minimum l_p-norm, where $p = 0, ½, 1, 2$ and ∞. Only For $0 \leq p \leq 1$, the solutions are sparse. On the other hand, only for $p \geq 1$ the corresponding optimization problem is convex.

distributed standard normal variables [Boche et al., 2015, pp. 13-17]:

$$\boldsymbol{\Phi} = \frac{1}{\sqrt{M}} \begin{bmatrix} \phi_{1,1} & \cdots & \phi_{1,N} \\ \vdots & \ddots & \vdots \\ \phi_{M,1} & \cdots & \phi_{M,N} \end{bmatrix} \tag{69}$$

Note that the scaling factor in Eq. (69) makes the columns of $\boldsymbol{\Phi}$ take l_2-norms of less than or equal to 1. This finding is relevant for the content of this thesis, as the matrices constructed in Chapters 4 and 5 are of similar nature to that of the matrix shown in Eq. (69).

2.3.5 Noise-aware l_1-norm Minimization

Since mathematical models usually do not perfectly mirror real-world systems and real measurements are generally corrupted by noise, the ideas described above have been extended to the following noise-aware version of Eq. (66):

$$\boldsymbol{y} = \boldsymbol{\Phi x} + \boldsymbol{e} \tag{70}$$

In Eq. (70), the vector $\boldsymbol{e} \in \mathbb{R}^M$ represents additive deterministic or random noise. Given this setting, sparse solutions can be found via a number of self-similar problem formulations some of which will be stated below. The problem most similar to the one stated in Eq. (68) is the *quadratically constrained basis pursuit* (QCBP) (also known as *noise-aware l_1-norm minimization*), which consists of solving the following minimization problem:

$$\min_{\boldsymbol{x} \in \mathbb{R}^N} \|\boldsymbol{x}\|_1 \quad \text{s.t.} \quad \|\boldsymbol{\Phi x} - \boldsymbol{y}\|_2 \leq \eta \tag{71}$$

In Eq. (71), the parameter $\eta \geq 0$ corresponds to the system noise level. Note that the standard basis pursuit problem stated in Eq. (68) is a special case of Eq. (71) where $\eta = 0$. The NSP has been adapted to this noise-aware setup, resulting in the so called l_2-*robust* NSP [Candès, 2008]. Candès showed that the RIP is also sufficient to guarantee the l_2-robust NSP, yielding the equivalence of the solutions of Eq. (71) and those of its l_0-counterpart under sightly stricter assumptions.

Similarly to the QCBP, sparse solutions to Eq. (70) can be obtained via the *Least Absolute*

Shrinkage and Selection Operator (LASSO) [Tibshirani, 1996]:

$$\min_{\boldsymbol{x}\in\mathbb{R}^N} \|\boldsymbol{\Phi x}-\boldsymbol{y}\|_2 \quad \text{s.t.} \quad \|\boldsymbol{x}\|_1 \leq \tau \tag{72}$$

with $\tau \geq 0$ being some parameter.

Yet another popular approach is to include the constraint as a regularization term in the objective function. This unconstrained version is referred to as, e.g., *l_1-regularized least squares problem* [Peter, 2016, p. 23] or *basis pursuit denoising* (BPDN) [Foucart and Rauhut, 2013, p. 64]:

$$\min_{\boldsymbol{x}\in\mathbb{R}^N} \lambda\|\boldsymbol{x}\|_1 + \|\boldsymbol{\Phi x}-\boldsymbol{y}\|_2^2 \tag{73}$$

All three above-stated formulations are strongly inter-linked as was brought to the point by Foucart and Rauhut [2013, Proposition 3.2.]:

(1) If $\boldsymbol{x}$ is a unique solution of the QCBP with $\eta \geq 0$, then there exists $\tau \geq 0$, such that $\boldsymbol{x}$ is a unique solution of the LASSO.
(2) If $\boldsymbol{x}$ is a solution of the LASSO with $\tau > 0$, then there exists a $\lambda \geq 0$ such that $\boldsymbol{x}$ is a solution of the BPDN.
(3) If $\boldsymbol{x}$ is a solution of the BPDN with $\lambda > 0$, then there exists an $\eta \geq 0$ such that $\boldsymbol{x}$ is a solution of the QCBP.

The values of λ, τ and η which make these problems equivalent are generally unknown a priori. However, since the optimum individual values of these parameters are generally unknown as well, one may use any of those three problem formulations for sparse signal recovery modelling. What *does* depend on the choice of the problem formulation is the *solving algorithm*, as there are different algorithms specifically designed to efficiently solve these problems individually. In this thesis, the BPDN was selected. More precisely, an extension of the BPDN to the more general joint sparsity model is used, which will be introduced in the following subsection. Moreover, the corresponding selected algorithm, FISTA, will be briefly described in Section 2.3.7.

2.3.6 Joint Sparsity

In some applications, including the multiresolution image fusion subjects this thesis concentrates on, the signals to be recovered are not only known to be sparse, but there is a priori information available about *structured sparsity* in groups of signals. One of the most prominent and well-studied types of structured sparsity is the *joint sparsity*: A group of signals, $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_D$, $D \in \mathbb{N}$, are called *jointly sparse*, if they are sparse on largely identical supports, i.e. $\text{supp}(\boldsymbol{x}_1) \approx \text{supp}(\boldsymbol{x}_2) \approx \ldots \approx \text{supp}(\boldsymbol{x}_D)$. Jointly sparse signals occur, for instance, when a group of measurement vectors are *mutually correlated*. Considering the potentially high correlation between adjacent bands in multispectral and – even more so – hyperspectral imagery (compare Section 2.1.6), the joint sparsity assumption will turn out to be excellently applicable to the image fusion problems at hand.

Instead of recovering the signals $x_1,\ldots, x_D$ individually, hoping that the expected joint sparsity pattern will naturally result from the given system and measurements, the signals are usually recovered *simultaneously* via a suitable modification of one of the above-stated single-signal minimization problems, i.e. Eqs. (71) to (73). The generalization of the BPDN, for instance, to the simultaneous recovery of jointly sparse signals using multiple measure-

ment vectors, is as follows:

$$\min_{\boldsymbol{X}\in\mathbb{R}^{N\times D}} \lambda\|\boldsymbol{X}\|_{1,2} + \|\boldsymbol{\Phi}\boldsymbol{X} - \boldsymbol{Y}\|_F^2 \tag{74}$$

In Eq. (74), the matrices $\boldsymbol{X} = [\boldsymbol{x}_1,\ldots,\boldsymbol{x}_D]$ and $\boldsymbol{Y} = [\boldsymbol{y}_1,\ldots,\boldsymbol{y}_D]$ are respectively formed by the group of signals to be recovered and the corresponding group of measurement vectors. The $l_{1,2}$-norm penalty term promotes row-sparse solutions $\boldsymbol{X}$, i.e. jointly sparse signals $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_D$. Intuitively, exploiting the prior knowledge of joint sparsity in the modelling should enhance the robustness of the signal recovery process. In fact, what has been proven is that, in terms of theoretical guarantees, under similar yet not significantly milder conditions on $\boldsymbol{\Phi}$ and s/N, algorithm extensions from individual to joint sparsity can be used to recover $\boldsymbol{X}$ exactly [Chen and Huo, 2006; Baraniuk et al., 2010]. In practice, however, joint sparse recovery may be more useful, as was analyzed by Eldar and Rauhut [2010]; Based on empirical testing, the authors showed that, in the average case, joint signal reconstruction requires fewer measurements (by means of a smaller dimension M of the measurement vectors $\boldsymbol{y}_d \in \mathbb{R}^M$) than individual signal reconstruction. In the same article, the authors showed that the failure probability decays exponentially in the number of channels D. Hence, an improvement of the reconstruction robustness can be expected, as long as the assumption of a joint sparsity model is reasonable for the application at hand.

Joint sparse signal recovery can not only improve robustness, but also the computational performance of the solving algorithm. The following section introduces recovery algorithms relevant for this thesis.

A more comprehensive overview on joint sparsity can be found in [Eldar and Kutyniok, 2012, Sections 1.7, 2.2 and 2.3]. There are related subjects, such as the more general block- (also known as group-) sparsity and other types of structured sparsity, which have been investigated along with the comparatively simple joint sparsity setting over the past decade. The following literature is recommended to read for an in-depth understanding of different types of structured sparsity: [Davies and Eldar, 2012], [Tropp, 2006], [Tropp et al., 2006], [Gribonval et al., 2008], [Fornasier and Rauhut, 2008], [Eldar and Mishali, 2009], [Eldar and Rauhut, 2010]. A general structured sparsity setting is discussed in the work on *model-based compressive sensing* by Baraniuk et al. [2010].

2.3.7 Recovery Algorithms

Once the minimization problem for sparse signal recovery (e.g. QCBP, LASSO, BPDN or a corresponding joint sparsity model) is chosen, the next step is to *solve* the selected minimization problem. There are different classes of algorithms developed for different types of minimization problems. For the sake of compactness, this section only discusses the class of algorithms that was used to solve the finite-dimensional BPDN-based problems occurring in this thesis, namely *iterative (soft) thresholding algorithms*.

Iterative thresholding Iterative thresholding algorithms (ITAs) are designed to solve the following type of unconstrained minimization problems, which are regularized by some separable lower semi-continuous function $\psi : \mathbb{R}^N \to \mathbb{R}$:

$$\min_{\boldsymbol{x}\in\mathbb{R}^N} \lambda\psi(\boldsymbol{x}) + \|\boldsymbol{\Phi}\boldsymbol{x} - \boldsymbol{y}\|_2^2 \tag{75}$$

The basic idea of iterative thresholding is to alternately perform a forward gradient descent step (also called *Landweber step*) on the second term in Eq. (75) and a backward gradient

descent step (also called *proximal step*) which involves a thresholding operation. If $t^{(j)}$ denotes the possibly adaptive stepsize in the j-th iteration, one full iteration step (also called *thresholded Landweber iteration*) is given by

$$\boldsymbol{x}^{(j+1)} = \mathcal{T}_\alpha\Big(\boldsymbol{x}^{(j)} + t^{(j)}\boldsymbol{\Phi}^T(\boldsymbol{\Phi}\boldsymbol{x}^{(j)} - \boldsymbol{y}), \psi\Big), \tag{76}$$

where $\mathcal{T}_\alpha$ denotes the proximity (thresholding) operator, defined as

$$\mathcal{T}_\alpha(\boldsymbol{x}, \psi) = \arg\min_{\boldsymbol{z}\in\mathbb{R}^N} \alpha\psi(\boldsymbol{z}) + \|\boldsymbol{z} - \boldsymbol{x}\|_2^2, \tag{77}$$

and $\alpha = \alpha(\lambda, t^{(j)}) \geq 0$ may vary from one ITA to another.

Iterative soft thresholding Two popular ITAs are the *iterative hard thresholding algorithm* (IHTA) and the *iterative soft thresholding algorithm* (ISTA), which describe the special cases of $\psi(\cdot) = \|\cdot\|_0$ and $\psi(\cdot) = \|\cdot\|_1$, respectively. Since only l_1-based optimization will be used in this thesis, IHTA will be no further extended on here. Corresponding literature can be found in, e.g., [Blumensath and Davies, 2009; Blumensath, 2012], as well as [Fornasier, 2010, Section 3.1.4].

For $\phi(\cdot) = \|\cdot\|_1$, it was shown in [Daubechies et al., 2004] that the proximity operator is uniquely determined by the *soft thresholding operator* $\mathcal{S}_\alpha(\cdot)$, which is defined, componentwise for $n = 1, \ldots, N$, by

$$\Big(\mathcal{S}_\alpha(\boldsymbol{x})\Big)_n = \begin{cases} \Big(1 - \frac{\alpha}{2|x_n|}\Big)x_n, & \text{if } |x_n| > \frac{\alpha}{2} \\ 0 & \text{otherwise} \end{cases}. \tag{78}$$

Given this operator, the ISTA is stated, for a fixed stepsize $t \leq 1$ and $\alpha = t\lambda$, in Algorithm 1, as was proposed independently by Starck et al. [2003] and Figueiredo and Nowak [2003] in different contexts.

Algorithm 1 Iterative Soft Thresholding Algorithm (ISTA)

Input: Stepsize $0 < t \leq 1$, regularization parameter $0 < \lambda$
Initilize: Set initial value for $\boldsymbol{x}$, e.g. $\boldsymbol{x}^{(0)} = \boldsymbol{0} \in \mathbb{R}^N$, $j = 0$
1: **while** Stopping criterion is not met **do**
2: $\quad \boldsymbol{x}^{(j+1)} = \mathcal{S}_{t\lambda}\Big(\boldsymbol{x}^j + t\boldsymbol{\Phi}^T\Big(\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}^j\Big)\Big)$ ▷ $\mathcal{S}_{t\lambda}(\cdot)$ as defined in Eq. (78)
3: $\quad j = j + 1$
4: **end while**
Output: Solution $\boldsymbol{x}$ of Eq. (75) with $\psi(\cdot) = \|\cdot\|_1$

ISTA is guaranteed to converge only for $\|\boldsymbol{\Phi}\|_2 < \sqrt{2}$ (see e.g. [Combettes and Wajs, 2005]). However, this is not a restriction on $\boldsymbol{\Phi}$, as it can always be met by rescaling $\boldsymbol{\Phi}$, λ and $\boldsymbol{y}$ in Eq. (73) accordingly without changing the actual problem and, hence, the solution $\boldsymbol{x}$.

Fast iterative soft thresholding The obvious advantage of ISTA lies in its simplicity. This simplicity, however, comes at the expense of slow convergence, which is why ISTA has been the subject of many means of acceleration in the past. Well-known techniques include iterative adaption of both the stepsize $t = t^{(j)}$ [see e.g. van den Berg and Friedlander, 2008] and regularization parameter $\lambda = \lambda^{(j)}$ [Dahlke et al., 2012]. One of the most established accelerators today, however, is the *fast iterative soft thresholding algorithm* (FISTA) [Beck and Teboulle, 2009], which is stated in Algorithm 2 and used in this thesis.

Algorithm 2 Fast Iterative Soft Thresholding Algorithm (FISTA)

Input: Stepsize $0 < t \leq 1$, regularization parameter $0 < \lambda$
Initilize: Set initial value for $\boldsymbol{x}$, e.g. $\boldsymbol{x}^{(0)} = \mathbf{0} \in \mathbb{R}^N$, $\boldsymbol{u}^{(0)} = \boldsymbol{x}^{(0)}$, $s^{(0)} = 0$, $j = 0$
1: **while** Stopping criterion is not met **do**
2: $\quad \boldsymbol{u}^{(j+1)} = \mathcal{S}_{t\lambda}\left(\boldsymbol{x}^{(j)} + t\boldsymbol{\Phi}^T\left(\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}^{(j)}\right)\right)$ ▷ $\mathcal{S}_{t\lambda}(\cdot)$ as defined in Eq. (78)
3: $\quad s^{(j+1)} = \frac{1+\sqrt{1+4(s^{(j)})^2}}{2}$
4: $\quad w^{(j)} = 1 + \frac{s^{(j)}-1}{s^{(j+1)}}$
5: $\quad \boldsymbol{x}^{(j+1)} = w^{(j)}\boldsymbol{u}^{(j+1)} + \left(1 - w^{(j)}\right)\boldsymbol{u}^{(j)}$
6: $\quad j = j + 1$
7: **end while**
Output: Solution $\boldsymbol{x}$ of Eq. (75) with $\psi(\cdot) = \|\cdot\|_1$

FISTA is still simple yet significantly faster than ISTA in that, FISTA achieves a worst-case convergence rate of $\mathcal{O}(n^{-2})$ while ISTA converges only linearly [Bredies and Lorenz, 2008]. Moreover, FISTA can be potentially further accelerated by e.g. parallelization of Algorithm 2, as was realized in the *parallel FISTA* (P-FISTA) [Peng et al., 2013], and backtracking strategies some of which were already proposed by Beck and Teboulle [2009] along with FISTA. However, those strategies only fructify for large-scale optimization problems and may even decelerate the algorithm for small-to-medium size problems. During the research work towards this thesis, several acceleration techniques have been tested, yielding the conclusion that the sequential FISTA as stated above is best suited for the specific image fusion models developed in Chapters 4 and 5. The best mean of acceleration, which is implemented in the J-SparseFI pan-sharpening method described in Chapter 4, is a better estimate of $\boldsymbol{x}^{(0)}$, which is possible thanks to knowledge about both the position and an estimate of the amplitude of the most significant non-trivial coefficient in $\boldsymbol{x}$.

Adaptation to joint sparsity Apart from its simplicity and good worst-case convergence properties, FISTA has the advantage of being easily modifyable for joint sparse recovery problems of the form shown in Eq. (74). Therefore, the vector-valued quantities $\boldsymbol{x}$, $\boldsymbol{u}$ and $\boldsymbol{y}$ in Algorithm 2 are substituted by corresponding matrices $\boldsymbol{X}$, $\boldsymbol{U}$ and $\boldsymbol{Y}$ each of which is composed of D columns, and the operator $\mathcal{S}_\alpha(\cdot)$ defined in Eq. (78) and used in Line 2 of Algorithm 2 is replaced by $\mathcal{J}_\alpha(\cdot)$, which is similarly defined row-wise as

$$\left(\mathcal{J}_\alpha(\boldsymbol{X})\right)_n = \begin{cases} \left(1 - \frac{\alpha}{2\|\mathbf{x}_n^T\|_2}\right)\mathbf{x}_n, & \text{if } \|\mathbf{x}_n^T\|_2 > \frac{\alpha}{2} \\ 0 & \text{otherwise} \end{cases}. \tag{79}$$

Stopping criterion What is still missing in order to be able to implement (F)ISTA is a suitable criterion that terminates the iteration. Several solutions have been proposed in the past some of which were compared in [Peter, 2016, Section 4.2.4] in a slightly more general (infinite dimensional) scope. The finding was that a penalization of the first order optimality conditions on the operator given by the minimizer in Eq. (74) is well suited as stopping criterion, because, in contrast to the other candidate criteria under comparison, it is strongly linked to the quality of the iterate, and it does neither depend on the solution $\boldsymbol{X}^*$ of Eq. (74), nor on the previous iterate. Specifically, the normalized stopping criterion used in this thesis

is determined by a residual term $r^{(j)}$ and a predefined tolerance ϵ_r and is given by

$$r^{(j)} := \frac{\|\boldsymbol{\xi}^{(j)}\|_1}{\|2\boldsymbol{\Phi}^T\boldsymbol{Y}\|_{1,2}} < \epsilon_r, \tag{80}$$

where $\boldsymbol{\xi}^{(j)}$ is defined component-wise for $n = 1,\dots,N$ as follows:

$$\xi_n^{(j)} = \begin{cases} \left\| \left(2\boldsymbol{\Phi}^T\left(\boldsymbol{\Phi}\boldsymbol{X}^{(j)} - \boldsymbol{Y}\right)\right)_n^T + \lambda \frac{(\mathbf{x}_n^{(j)})^T}{\|(\mathbf{x}_n^{(j)})^T\|_2} \right\|_2, & \text{if } \|(\mathbf{x}_n^{(j)})^T\|_2 > 0 \\ \max\left\{0, \left\| \left(2\boldsymbol{\Phi}^T(\boldsymbol{\Phi}\boldsymbol{X}^{(j)} - \boldsymbol{Y})\right)_n^T \right\|_2 - \lambda \right\} & \text{otherwise} \end{cases} \tag{81}$$

Note that Eqs. (80) and (81) conform to both the joint sparsity extension of (F)ISTA and the original vector-valued version. Moreover, the multiplication $[\boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{X}^{(j)}]$ does not imply additional computational costs, as it is re-used in (F)ISTA.

As outlined in the introduction of this section, the field of sparse recovery is vast. There are a number of books on this topic, which the interested reader is recommended to study for more in-depth information: [Fornasier, 2010; Elad, 2010; Eldar and Kutyniok, 2012; Foucart and Rauhut, 2013].

3

State of the Art in Remote Sensing Data Fusion

The fusion of data acquired by complementary sensors is the key to producing high-quality remote sensing products beyond the physical limits associated with individual sensors. One of the most active research domains in remote sensing data fusion addresses the scenario in which the complementary sensors feature different spatial and spectral resolutions. Recall Section 2.2 for an introduction of that subject. This chapter provides an overview of the various categories in remote sensing multiresolution data fusion, summarizes the state-of-the-art and outlines the main contributions of this thesis. In particular, Sections 3.1 and 3.2 discuss existing pan-sharpening and hyperspectral resolution enhancement techniques, respectively. Section 3.3 is dedicated to the young field of sparse representation-based image fusion. The contributions of this thesis are outlined and brought into context in Section 3.4.

3.1 Pan-Sharpening

Given the diversity of former, existing and planned single-platform multispectral (MS) and panchromatic (Pan) sensor pairs on the one hand, and the undoubted public interest in high-resolution Earth observation data on the other hand, researchers have spent enormous efforts to develop solutions to the pan-sharpening problem in the past decades. As pan-sharpening can be considered an ill-posed inverse problem, where the degrees of freedom increase with both the GSD ratio and the difference in numbers of bands between the input images, finding a unique solution – i.e. high resolution fusion product – may require strong assumptions. The more information about the data acquisition process, atmospheric conditions, instrument design and the observed scene is known and incorporated in the fusion model, the higher the chance of producing high resolution imagery which is close to the high-resolution multispectral "truth".

Before summarizing the existing literature on that subject, it is useful to indicate what characterizes a "good" pan-sharpening method:

(1) Homogeneous areas at the spatial scale larger than one low-resolution pixel should be recovered *exactly*.

(2) The fusion product should be visually sharp. That is, details, such as lines, should be spatially separable at the resolution level of the Pan sensor.
(3) Spectral profiles should be well-preserved at the majority of pixels in the scene. In other words, the fusion scheme should not distort colors while enhancing the spatial resolution of the MS input image.
(4) The high-resolution ground information acquired by the Pan sensor collectively in a broad wavelength interval should be split – i.e. redistributed – into smaller wavelength ranges by means of those MS channels whose SRFs are covered by that of the Pan sensor. For instance, if the Pan sensor is responsive in both the blue and the red spectral portions, and senses possibly adjacent red and blue objects, those objects should be neatly discriminated into blue and red details in the high-resolution MS fusion product.
(5) Some Earth observing MS imagers, such as WorldView-2/3 and Landsat-8/OLI (see Tab. 1), comprise channels that are not covered by the spectral response of the Pan sensor. Those MS bands may contain details that are not visible in the Pan image. A good pan-sharpening technique should be principally able to recover such details.

The following literature review reveals that the majority of existing methods categorically fail to meet one or multiple of these quality indicators.

Comprehensive review articles on existing pan-sharpening techniques compiled at various points in time show the development in this field over the past three decades [Ehlers, 1991; Pohl and Genderen, 1998; Wang et al., 2005; Alparone et al., 2007; Thomas et al., 2008; Amro et al., 2011; Jawak and Luis, 2013; Vivone et al., 2015]. There are different ways of categorizing pan-sharpening methods as was discussed in detail recently [Pohl and van Genderen, 2015]. Classification is usually based on the key element or mathematical operation used in the fusion procedure. Two well-distinguishable and widely accepted categories describe those algorithms based on *component substitution* (CS) and *multiresolution analysis* (MRA), respectively. Other categories include *modulation*-based fusion [Jawak and Luis, 2013; Basaeed et al., 2013; Pohl and van Genderen, 2015] (also known as fusion based on *relative spectral contribution* [Amro et al., 2011]) and fusion based on *statistics of the image* [Amro et al., 2011; Pohl and van Genderen, 2015]. Moreover, several *hybrid methods* have been proposed, which combine approaches from multiple categories. In the latest review article on pan-sharpening, which is also one of the most comprehensive ones to date and recommended for a deeper look into this subject, Vivone et al. [2015] showed that many of the algorithms that are often categorized as modulation- or statistics-based, can be considered as either CS- or MRA-like methods as well. The following paragraphs follow this compact classification scheme for simplicity. *Sparse representation*-based image fusion is dedicated its own category, which will be discussed separately in Section 3.3.

Component Substitution Component substitution (CS) methods, also referred to as *spectral methods* [Aiazzi et al., 2017] or *projection substitution methods* [Thomas et al., 2008], involve spectral transformations, which are supposed to separate the spatial from the spectral information in different components. CS-based fusion is conducted on the spatial grid of the Pan image, which requires an initial interpolation and upsampling of the MS image. Subsequently, the upsampled MS data is projected into a feature space via some – typically linear – spectral transformation. Then, histogram matching is performed to adapt the intensity range of the Pan image to the intensity range of the component (i.e. feature band) supposedly holding the spatial information of the MS image. Thereafter, that component is substituted by the histogram-matched Pan image. Finally, the modified data is projected back into the original color space.

One of the advantages of CS is its low computational complexity. As was first pointed out

by Tu et al. [2001] and further generalized by Dou et al. [2007]; Aiazzi et al. [2007]; Vivone et al. [2015], CS-based methods that use a linear transformation and substitute only a single component have a common representation which does not require an explicit calculation of the forward and inverse transformation. If $\boldsymbol{X}$ and $\tilde{\boldsymbol{Y}}_k$, $1 \leq k \leq N$, denote the histogram-matched Pan and the k-th band of the interpolated MS image, respectively, any CS-based technique can be expressed as follows:

$$\hat{\boldsymbol{Y}}_k = \tilde{\boldsymbol{Y}}_k + g_k \left(\boldsymbol{X} - \sum_{l=1}^{N} \omega_l \tilde{\boldsymbol{Y}}_l \right), \quad k = 1, \ldots, N \tag{82}$$

In Eq. (82), g_k denotes a gain and ω_k a weight for the k-th MS channel. Considering this representation, CS-based fusion techniques only differ in $\boldsymbol{g} = [g_1, \ldots, g_N]$ and $\boldsymbol{\omega} = [\omega_1, \ldots, \omega_N]$ which, in turn, primarily depend on the spectral transformation. The spectral transformations most commonly used for CS-based pan-sharpening include *intensity-hue-saturation* (IHS) [see e.g. Carper et al., 1990; Tu et al., 2004; Rahmani et al., 2010], *principal component analysis* (PCA) [see e.g. Chavez et al., 1991; Shah et al., 2008], *Gram Schmidt* (GS) [see e.g. Laben and Brower, 2000; Aiazzi et al., 2007, 2009] and the *Brovey transformation* (BT) [see e.g. Gillespie et al., 1987; Zhang, 1999; Tu et al., 2001]. Explicit derivations of the gains and weights for those and other transformations can be found in [Vivone et al., 2015]. However, it does not require further specifications to notice some general properties of this approach with respect to the quality indices listed above: (1) Perfectly homogeneous areas can indeed be exactly recovered if, e.g., $\boldsymbol{g} = \boldsymbol{0}$. Otherwise, the Pan channel will introduce not only details but systematic global spectral distortion; (2) If g_k is sufficiently large and the ω_ls sufficiently small, spatial details that are visible in the Pan image will be transferred into $\hat{\boldsymbol{Y}}_k$ and, hence, preserved. However, apart from the Pan details, also the interpolated MS image is globally added to $\hat{\boldsymbol{Y}}_k$. Thus, the fusion results always contains a blurred component. (4) Discriminating spectrally distinct details, that are equally present as gray values in the Pan image, is principally possible. However, for diverse scenes, a spatially adaptive calculation of $\boldsymbol{g}$ – as is done e.g. by many MRA-based methods listed in the following paragraph – is indispensable. (5) This criterion cannot be met in general; If an MS channel contains details that are not visible in the Pan image *at exactly the same location*, those details cannot be recovered by CS-based fusion methods.

Multiresolution Analysis While the idea of CS is to use *spectral* transformations of the MS data to approximate the low-frequency component of the Pan image and, thus, extract its high-frequency information, multiresolution analysis (MRA) involves *spatial* transformations of the Pan image to obtain its high-frequency information. Hence, MRA-based methods are also referred to as *spatial methods* [Aiazzi et al., 2017]. If $\boldsymbol{X}_L$ denotes a spatially degraded version of the Pan image $\boldsymbol{X}$, $\boldsymbol{G}_k$ is a gain matrix of the same dimension as $\boldsymbol{X}$ and the operator "$\circ$" denotes the Hadamard product, then MRA-based pan-sharpening methods can be written band-wise as

$$\hat{\boldsymbol{Y}}_k = \tilde{\boldsymbol{Y}}_k + \boldsymbol{G}_k \circ (\boldsymbol{X} - \boldsymbol{X}_L), \quad k = 1, \ldots, N. \tag{83}$$

This general paradigm has been formulated first by Ranchin and Wald [2000], who named it the ARSIS concept (from the French *Amélioration de la Résolution Spatiale par Injection de Structures*, which stands for *spatial resolution enhancement by structure injection*). Pan-sharpening techniques that can be cast into Eq. (83) are uniquely defined by the low-pass filtered image $\boldsymbol{X}_L$ and the injection gains $\boldsymbol{G}_k$, both of which are determined by the underlying MRA technique. Essentially, $\boldsymbol{X}_L$ is obtained through a recursive decomposition scheme, i.e. MRA, which is used to generate a sequence of 2-D signals of progressively reduced spa-

tial information by a repeated application of low-pass filters. Typical MRA techniques used for pan-sharpening are based on wavelets [see e.g. Garguet-Duport et al., 1996; Zhou et al., 1998; Otazu et al., 2005], curevelets [see e.g. Choi et al., 2005; Nencini et al., 2007], contourlets [see e.g. Lillo-Saavedra and Gonzalo, 2007; Yang and Jiao, 2008] and Laplacian pyramids [see e.g. Aiazzi et al., 1999, 2002, 2006]. Note that, in contrast to CS-based methods, the injection gains derived from MRA may differ locally. Specific dependencies of $\boldsymbol{G}_k$ and $\boldsymbol{X}_L$ on the inputs $\boldsymbol{Y}_k$ and $\boldsymbol{X}$, as well as on possible additional parameters for different MRA schemes can be found in [Vivone et al., 2015]. Among the various MRA-based fusion methods, the *additive wavelet luminance proportional* (AWLP) method by Otazu et al. [2005] is specifically worth-mentioning, as – given its simplicity and consistently competitive performance – it has become one of the benchmark methods used for fusion quality comparison. It is based on the *á trous* wavelet transform and was developed as an improved version of the similarly established *additive wavelet* (AWL)-based pan-sharpening method proposed by Nunez et al. [1999]. Improved spectral information preservation is achieved through a proportional injection of the high-pass Pan details to the low-pass MS components. In the GRSS data fusion context of 2006, AWLP achieved rank two out of eight pan-sharpening methods under comparison [Alparone et al., 2007], while, in a more recent study on the applicability of pan-sharpening methods to DubaiSat-1 data, AWLP even outperformed all of the eight other algorithms under comparison [Basaeed et al., 2013]. Hence, among others, AWLP was selected to demonstrate the performance of the joint sparse representation-based J-SparseFI method proposed in Chapter 4.

Considering the similarity between the CS- and MRA-based fusion frameworks expressed by Eqs. (82) and (83), the general comments on the MRA approach with respect to the five above-listed quality indicators are nearly identical to those made for the CS-approach. The main advantage of MRA over CS is the typically spatially varying gain factor, which may lead to improvements at least in terms of indicators (1), (2) and (4). Quality indicator number (5) remains an open issue, which cannot be resolved via MRA-based image fusion either. In order to be able to recover spatial details in the MS fusion product which are not observable by the Pan sensor, non-local detail injection is required. That is, objects that are visible in those MS bands are not correlated with the Pan image may require local reconstruction based on high-resolution information stored in a database such as a dictionary. This is one of the ideas of the young field of sparse representation-based image fusion which will be discussed in the Section 3.3.

3.2 Hyperspectral-Multispectral Data Fusion

The spatial resolution enhancement of hyperspectral (HS) data via data fusion using auxiliary higher-resolution MS imagery does not have a historical development as diverse and long as pan-sharpening. Existing methods usually incorporate multiple theories, which complicates categorization. Recently, a comprehensive review article on HS-MS image fusion was published in IEEE Geoscience and Remote Sensing Magazine, which the interested reader is recommended to refer to for a comprehensive overview of the subject [Yokoya, Grohnfeldt, and Chanussot, 2017]. This section roughly follows the structure proposed in that work.

Despite the considerable overlaps between potential categories, there are a few methodological elements that may be used for the grouping of algorithms. One of those elements is the splitting of the HS-MS fusion problem into various sub-problems each of which can be

solved using a conventional pan-sharpening method. This category can be regarded as *generalization of pan-sharpening techniques* or *pan-sharpening-based HS-MS image fusion*. Fusion techniques not belonging to this category are specifically designed for HS-MS fusion and do not call an existing pan-sharpening method. Considering the relatively high correlation between adjacent HS bands, many HS-MS fusion methods conduct the fusion procedure on a linear spectral subspace – hence the category of *subspace-based methods*. Moreover, the recent activities in the field of spectral unmixing in the hyperspectral signal processing community has led to the development of *spectral unmixing-based HS-MS image fusion methods*. As was mentioned above, some methods may incorporate multiple of these techniques, which is why the remainder of this discussion is not strictly divided by category. Only those methods based on *sparse representations* are separately discussed in Section 3.3.

As pan-sharpening can be considered a special case of the HS-MS fusion problem, efforts have been made towards generalizing existing pan-sharpening methods for HS-MS fusion. One of the first attempts of pan-sharpening-based HS-MS fusion was presented by Gomez et al. [2001], who proposed to sharpen one HS channel at a time using the MS band which corresponds to the same wavelength range. The authors tested their idea on an MRA-based wavelet technique. As was discussed later by Zhang and He [2007], the performance was highly sensitive to miss-association between HS and MS bands. Consequently, visible spectral distortions lowered the usefulness of the resulting high-resolution HS data for subsequent analyses. More recently, in [Grohnfeldt et al., 2013], a similar technique was proposed, but instead of sharpening all HS band individually, *groups* of adjacent HS channels were identified according to the SRFs of the MS sensor. The HS channels belonging to the same group were processed *jointly*. Shortly afterwards, Chen et al. [2014] proposed a similar framework, differing in that, HS bands corresponding to spectral regions that are not covered by the MS SRFs are pan-sharpened by an MS band that is arithmetically synthesized from spectrally adjacent MS bands. The main drawbacks, all of the afore-mentioned approaches implicate, are as follows: (1) In the proposed schemes, the grouping required manual interference; (2) Different HS groups correspond to non-overlapping spectral portions which may lead to discontinuities in the spectral profiles of the HS fusion product. Those two issues were tackled by Grohnfeldt et al. [2014]; Grohnfeldt and Zhu [2015a], who developed an automatic *spectral grouping concept* based on the sensors' spectral responsivity. In that approach, the HS groups are allowed to spectrally overlap. In a direct comparison to Chen et al.'s method, spectral information preservation capabilities were improved visibly and quantitatively [Grohnfeldt and Zhu, 2015a]. Yet another means of generalization was proposed by Selva et al. [2015] who developed a framework called *hypersharpening*. The idea is to synthesize – for each HS band – a high-resolution image as a linear combination of MS band images via linear regression. The authors demonstrated the effectiveness of the proposed scheme using an MRA-based pan-sharpening technique. It was shown that the synthesized high-resolution bands produced by hypersharpening could lead to significantly better fusion results than a similar yet simpler approach in which – for each HS band – one high-resolution MS band – namely the most correlated one – is *selected* from the available MS bands rather than synthesized.

The obvious advantage of the above-discussed frameworks, each of which is proposed to generalize pan-sharpening methods for the fusion of HS and MS imagery, is that they can be combined with any existing or future pan-sharpening method. Furthermore, they do not implicate large implementation or computational overheads and they can be tested quickly with publicly available pan-sharpening source codes. On the other hand, pan-sharpening methods are primarily designed to *spatially enhance* MS imagery. The key aspect of HS data, however, is the spectral information retrieval for material identification purposes. Hence,

the ideal HS-MS image fusion method must involve the simultaneous consideration of the full spectral dimension, which is why more sophisticated methods that are specifically designed for HS-MS fusion may be better suited for subsequent HS data analyses and applications.

One of the first methods specifically designed for HS-MS fusion was proposed by Eismann and Hardie [2004] and supplemented in [Hardie et al., 2004; Eismann and Hardie, 2005]. In their works, the authors exploit the inherent spectral characteristics of the scene by conducting the fusion process on a spectral subspace spanned by a set of basis vectors. The fusion model is formulated using a Bayesian paradigm combining a maximum *a posteriori* (MAP) estimation process with a stochastic mixing model (SMM). That fusion framework, which is commonly referred to as "MAP-SMM", was the first approach to use PCA for HS-MS image fusion and fully conduct the fusion process on a subspace – an idea which has been inspiring the development of follow-up HS-MS fusion methods until today [Zhang and He, 2009; Wei et al., 2015b; Simões et al., 2015].

One of the most active sub-domains in HS-MS image fusion to date is based on spectral unmixing, where the idea is to spatially enhance the HS data by identifying contributing material spectral signatures – i.e. endmembers – and corresponding abundances at a subpixel level. More precisely, the high-resolution HS image product is reconstructed as the product of a hyperspectral endmember matrix – also referred to as mixing matrix – and a high-resolution abundance matrix. Although the idea of using spectral unmixing for hyperspectral resolution enhancement was presented already around 2000 [Gross and Schott, 1998; Zhukov et al., 1999; Robinson et al., 2000], this field substantially revived around 2011 [Berné et al., 2010; Yokoya et al., 2011b,a; Bieniarz et al., 2011; Yokoya et al., 2012], which was when the field of hyperspectral unmixing began to receive increasing attention from the hyperspectral signal processing community. Berné et al. [2010] presented an HS-MS fusion method based on nonnegative matrix factorization (NMF) for mid-infrared astronomy. The low-resolution HS image is unmixed by NMF while the high-resolution abundance maps are obtained from the MS image via linear least squares approximation. The resampled endmember signatures are supposed to be known *a priori* and are not modified during the fusion process. The most well-established unmixing-based fusion algorithm today is the *coupled nonnegative matrix factorization* (CNMF)-based HS-MS data fusion scheme proposed by Yokoya et al. [2012]. The idea is to alternatingly decompose the HS and the MS data into HS and MS spectral endmember- and abundance matrices, respectively. Between the iterations, all four matrices are updated using the system's SRFs and PSFs. Information about initial values of the presumably pure endmembers is obtained solely from the input HS data via *vertex component analysis* (VCA) [Nascimento and Dias, 2005b]; VCA is based on the singular value decomposition (SVD) and finds a pre-defined number (typically less than the number of HS bands) of smooth spectra and can, thus, be used for subspace transformation alternatively to PCA or SVD. Another subspace-based method called *HySure* was proposed by Simões et al. [2015]. The authors adopted the SRF and PSF-based sensor observation that was used in, e.g., [Eismann and Hardie, 2005] and [Yokoya et al., 2012], but regularized the resulting optimization problem by total variation (TV) penalty terms. Furthermore, instead of approximating the PSFs of the HS sensor using Gaussian functions, they are estimated from the input data. The experimental results shown and discussed in Chapter 5 will confirm that the inherent denoising properties of HySure makes this algorithm one of the most competitive HS-MS fusion methods today. Another approach, which combines the spectral unmixing approach of CNMF with the TV penalization idea of HySure, is proposed in [Lanaras et al., 2015], where a projected gradient method is introduced into the alternate update process of the endmember and abundance matrices.

After MAP-SMM [Eismann and Hardie, 2004], Wei et al. [2015b] was the second group to formulate the HS-MS fusion problem as Bayesian estimation process. The authors use a minimum mean square error (MMSE) estimator instead of the MAP estimator used by Eismann and Hardie, which – among other modifications – led to an overall fusion quality improvement at the expense of computational simplicity. The latter drawback was mitigated by the same group shortly afterward by casting their previous Bayesian-based problem formulation into an explicitly solvable Sylvester equation [Wei et al., 2015c]. That new method was named *Fast fUsion based on Sylvester Equation* (FUSE). Similarly to MAP-SMM, HySure and other above-mentioned techniques, the fusion process implemented in FUSE is conducted on a (PCA-based) spectral subspace. If the intrinsic spectral dimension of the hyperspectral data, i.e., the dimension of the spectral subspace, is larger than the number of multispectral channels, the underlying Sylvelster equation has no unique solution. In this case, the authors proposed three means of regularization: (1) a naive Gaussian prior, (2) patch-based sparse representations, and (3) total variation. In a direct comparison, the authors experimentally found that the sparse representation-based regularization led to the best fusion results. This finding was confirmed in [Yokoya et al., 2017], where FUSE (with naive Gaussian prior) and FUSE-S (with sparse representation-based regularization) were both included in a comparative study: Yokoya et al. [2017] found that FUSE-S consistently outperforms FUSE in a large variety of multi-sensor fusion scenarios. Furthermore, compared to other recently proposed fusion techniques, FUSE-S performs particularly well when either the GSD ratio between the input sensors is large, such as eight, or the spectral range of the HS sensor is largely covered by MS bands. In more challenging scenarios, however, where VNIR-SWIR hyperspectral data is fused with VNIR-only multispectral imagery, FUSE-S was found to have representation difficulties. The latter issue and other properties of FUSE-S are further discussed in the review of sparse representation-based data fusion algorithms in the following section.

3.3 Sparse Representation Based Image Fusion

There is a high potential for improvement of the data fusion product if the fusion process is conducted locally rather than or in addition to globally. The diversity and heterogeneity of the Earth's surface indicate that a global derivation of physical parameters, that are then assumed to be valid for the full scene, can merely roughly approximate the true locally varying surface properties. This observation is one of the key motivations for conducting the image fusion process patch-wise. Moreover, the fact that both pan-sharpening and HS-MS image fusion generally lead to a severely ill-posed problem requires *a priori* system information to reduce the inherent degrees of freedom. The idea of sparse representation-based image fusion is to reconstruct the target high-spectral-high-spatial resolution image patch-wise subject to the assumption that any patch can be expressed as a linear combination of only a few of many available candidate patches. Formulated as a sparse recovery problem, patches are assumed to feature a sparse representation in an overcomplete dictionary composed of candidate patches.

Pan-Sharpening Sparse signal representation of image patches has been explored to solve the pan-sharpening problem in [Li and Yang, 2011] for the first time. In that work, MS image patches are assumed to have a sparse representation in a dictionary sampled from high-resolution MS images acquired by "comparable" sensors. The authors demonstrated that their method performs competitively to superiorly compared to some of the traditional methods discussed in Section 3.1. However, since the algorithm of Li and Yang requires

training images from a usually unavailable high-resolution MS sensor, which features not only a spatial resolution at least as high as the Pan instrument but also similar spectral characteristics to the MS sensor, its applicability is limited. For example, pan-sharpening of data with the highest resolution available is not possible per definition with this algorithm. To cope with that problem, a composition of a joint dictionary of oversampled low-resolution MS and high-resolution Pan images was proposed in [Jiang et al., 2012], where the high-resolution MS image is assumed to be sparse in that dictionary. Still, this method requires large collections of low-resolution MS and high-resolution Pan image pairs. In [Zhu and Bamler, 2013], the authors proposed a pan-sharpening method named *Sparse Fusion of Images* (SparseFI, pronounced "sparsify") that can be used in a broader application domain. In contrast to Li and Yang's method, SparseFI explores the sparse representation of MS image patches in a dictionary trained only from the Pan image at hand. Therefore, no high-resolution MS images from other sensors are required. It has been demonstrated that the SparseFI algorithm does not assume any spectral composition model of the Pan image and achieves robust performance against even gross spectral model errors. An extension of the SparseFI algorithm is proposed by Jiang et al. [2014], who implemented a two-step sparse coding concept.

Hyperspectral-Multispectral Image Fusion Sparse representations have been used for HS-MS image fusion via two approaches in the recent past: One is the patch-based image reconstruction, which is based on the same rationale as was discussed in the previous paragraph on sparse representation-based pan-sharpening, but with extension to HS-MS image fusion. The second approach falls into the category of spectral unmixing-based HS-MS image fusion, which was briefly discussed in Section 3.2. The main difference between *sparse* spectral unmixing and other spectral unmixing-based approaches such as CNMF is that the mixing matrix is overcomplete and the abundance vectors are assumed to be sparse. This assumption is physically meaningful if the underlying unmixing model uses a large spectral library that is composed of real material spectra. HS-MS data fusion algorithms based on sparse spectral unmixing have been developed by, e.g., Wycoff et al. [2013]; Akhtar et al. [2014] and Bieniarz et al. [2015]. Since this thesis focuses on exploiting sparse representations of *image patches* in dictionaries, the remainder of this paragraph discusses the development of this first branch rather than sparse spectral unmixing-based data fusion. Unless stated otherwise, the term "sparse representation-based image fusion" will refer to the sparse representation of *image patches* rather than spectral profiles in the remainder of this thesis.

There have been two research groups contributing to the field of sparse representation-based HS-MS image fusion in the past; one being the author of this dissertation et al. [Grohnfeldt et al., 2013, 2014, 2015; Grohnfeldt and Zhu, 2015b], the other group being Wei et al. [2014, 2015a].

The first approach towards solving the HS-MS image fusion problem based on sparse representations was published in [Grohnfeldt et al., 2013]. In that work, the authors presented a first attempt to generalize a sparse representation-based pan-sharpening method for solving the HS-MS data fusion task. In particular, the authors proposed to jointly sharpen groups of adjacent HS channels using a preliminary version [Zhu, Grohnfeldt, and Bamler, 2013a,b] of the *Jointly Sparse Fusion of Images* (J-SparseFI) pan-sharpening algorithm presented in Chapter 4. The HS channels are manually grouped according to the spectral overlap between the MS and HS bands. First results indicated that the joint sparse representation-based fusion of images is well suitable for the joint processing of adjacent HS channels and leads to stable spatial resolution enhancement. However, three drawbacks of this first approach can be identified: (1) the grouping of HS bands is performed manually, which is not practical as

it leaves the association of HS and MS bands to the user, (2) the HS groups are spectrally disjoint, which was shown to cause spectral discontinuities in the spectral profiles of the fusion product, and (3) no solution was provided as to how to sharpen HS channels whose spectral responses do not intersect with those of the MS instrument.

The first two issues were mitigated in [Grohnfeldt et al., 2014], where the authors proposed a *spectral grouping concept* (SGC) which automatically groups all HS bands into mutually overlapping bundles of adjacent channels in order to reduce the previously encountered spectral discontinuities. Each group is then associated and fused with possibly multiple high-resolution MS bands, depending on the spectral responsivity of both sensors. While SGC was initially designed to be combined with J-SparseFI, it was later demonstrated to work well also with conventional pan-sharpening methods [Grohnfeldt and Zhu, 2015a]. Still, HS bands outside of the spectral range of the MS imager remained an open issue.

This last drawback was resolved shortly afterwards by the same authors, who proposed a combined sparse representation and unmixing based hybrid hyperspectral resolution enhancement method [Grohnfeldt and Zhu, 2015b]. That approach uses the *joint sparsity model* (JSM) of J-SparseFI (see Chapter 4) to sharpen specific bundles of HS channels, namely those that are spectrally located around the centers of the MS SRFs, and use the resulting high-resolution HS bands as additional input to a modified version of CNMF [Yokoya et al., 2012]. This new spatial-spectral unmixing-based approach renders the ill-posed problem implicated in CNMF, i.e., the estimation of the high-resolution abundance matrix from multispectral measurements, well-posed by imposing the JSM-sharpened HS bands as additional measurements.

Meanwhile, Wei et al. [2014, 2015a] proposed a novel HS-MS fusion technique, which adopts the sensor observation model that is incorporated in, e.g., MAP-SMM [Eismann and Hardie, 2004], CNMF [Yokoya et al., 2012], Lanaras's method [Lanaras et al., 2015], and HySure [Simões et al., 2015], and uses sparse representations of image patches to regularize the resulting inverse problem. Similarly to Hardie et al. [2004] and Simões et al. [2015], the authors reduce the inherent degrees of freedom of the ill-posed problem by conducting the fusion process on a linear spectral subspace. Like Hardie et al. [2004], the authors use PCA for subspace transformation. Finally, sparse representations of image patches are utilized to regularize the resulting optimization problem. In particular, sparse representation-based patch reconstruction is performed, for each principal component band separately, using a global dictionary that is trained via online dictionary learning (ODL) [Mairal et al., 2009] from a rough estimate of the PCA-projected low-dimensional target hyperspectral image. In praxis, the optimization problem is solved for the PCA-projected low-dimensional target image $\boldsymbol{U}$ and the sparse patch coefficients $\boldsymbol{A}$ in an alternating manner. In that process, $\boldsymbol{U}$ is optimized for using the alternating direction method of multipliers (ADMM) [Boyd et al., 2011], whereas $\boldsymbol{A}$ is obtained via orthogonal matching pursuit (OMP) [Tropp and Gilbert, 2007] in the first iteration and least squares regression on the initially identified supports in subsequent iterations.

In a follow-up work [Wei et al., 2015c], the same authors proposed to re-formulate the optimization problem for $\boldsymbol{U}$ (with $\boldsymbol{A}$ fixed) as a Sylvester equation, which has an explicit closed-form solution and, thus, reduces the processing time implicated by ADMM. The authors showed that this new technique, called *Fast fUsion based on Sylvester Equation* (FUSE) coupled with sparse regularization (here and in [Yokoya et al., 2017] referred to as "FUSE-S"), outperforms the original sparse representation-based method [Wei et al., 2015a] in terms of computational efficiency while yielding comparable yet slightly better quality measures for the fusion products. Hence, FUSE-S is used – in addition to MAP-SMM, CNMF, Lanaras'

method, HySure, and other techniques – to represent the state of the art in HS-MS data fusion in the experimental part of Chapter 5.

3.4 Contribution of this Thesis

This thesis presents a new perspective on multiresolution data fusion. It demonstrates how the yet unexplored state-of-the-art tools of joint sparse signal recovery can be effectively used to reliably generate high-quality image fusion products. The primary contributions comprise solutions to the general multi-sensor multiresolution pan-sharpening and hyperspectral-multispecrtral data fusion problems. Sophisticated algorithms have been thoroughly designed, highly parallel software solutions have been implemented and comprehensive tests has been conducted, which demonstrate the effectiveness of the proposed data fusion schemes.

Joint Sparse Representation-based Pan-sharpening Although the recently proposed sparse reconstruction-based methods lead to promising results, at least four drawbacks of this type of image fusion algorithms can be pointed out, which carry potential for improvement.

(1) Compared with other conventional methods, they implicate a high level of computational complexity.
(2) Existing sparse reconstruction-based fusion methods use a single or coupled *global* dictionary for the reconstruction of *all* image patches. In particular for large-scale images, which comprise multiple thousands of patches, it is not possible to create a single coupled dictionary pair which is equally well suited for all individual patch reconstruction problems. Moreover, the number of atoms in the dictionaries has a major impact on the processing time, as will be discussed in Chapter 4.
(3) No consideration has been given to the possibility of mutually correlated information in different multispectral channels. Such correlations constitute a so-far-unexploited prior to the solution, the *joint sparsity* (see Section 2.3.6).
(4) Methods, whose coupled dictionary pair(s) are trained directly from the input data, such as SparseFI and its two-step extension, can be applied to a broader range of applications than previously proposed approaches; However, they also require the panchromatic and multispectral images to cover the same wavelength region, which is not a valid assumption for many modern sensors (see Tab. 1 on page 38).

In the scope of the work towards this dissertation, a sophisticated sparse image fusion algorithm was developed, which has been named *Jointly Sparse Fusion of Images* (J-SparseFI) algorithm and has been published in [Zhu, Grohnfeldt, and Bamler, 2016]. It is based on the SparseFI algorithm [Zhu and Bamler, 2013] and overcomes the aforementioned four drawbacks of the existing sparse pan-sharpening algorithms as is explained in the following.

(1) The developed solution to the problem related to the computational costs is two-fold: First, the computational time of each patch reconstruction problem is reduced by introducing a controllable dictionary size that is independent of the sizes of the input images. Second, a fully parallelizable framework is proposed and implemented. That is, the proposed scheme allows for each of the possibly overlapping patches to be processed independently instead of sequentially as was proposed in the previous works.
(2) A coupled *locally adjustive* dictionary pair is specifically generated for each patch reconstruction problem instead of one global dictionary possibly containing all candidate atoms.

(3) The proposed J-SparseFI algorithm exploits the signal structural correlation between multispectral channels. This is done by introducing a *joint sparsity model* (JSM). The JSM makes use of the distributed compressive sensing theory [Sarvotham et al., 2005; Baron et al., 2009] that restricts the solution of an underdetermined system by considering an ensemble of signals being jointly sparse.
(4) J-SparseFI offers a practical solution for the case of spectral range mismatches between a panchromatic and a multispectral instrument. One of the developed features is a module to conduct a *channel mutual correlation analysis*. This module uses the sensor-inherent spectral response information to identify fusion sub-problems: The multispectral channels are automatically assigned to one or more primary groups of joint channels, secondary groups of joint channels, and individual channels. The highly correlated primary groups of joint channels, whose spectral ranges match well with the one of the panchromatic image, are jointly sharpened via the JSM using a coupled dictionary pair built up from the panchromatic image. Subsequently, individual channels that suffer from sensor spectral mismatch are sharpened individually in a sequential manner by a SparseFI-like fusion scheme using updated dictionary pairs built up from previously reconstructed high-resolution channels. Finally, J-SparseFI jointly sharpens the secondary groups of via JSM using updated dictionary pairs built up from the reconstruction results obtained for the primary groups.

The proposed methodology is thoroughly validated in Sections 4.4 and 4.5. Detailed descriptions of all adjustment parameters, comprehensive analyses of their impact on the quality of the data fusion product and the processing time, and recommendations of default settings are provided. The quality of the data fusion products obtained via J-SparseFI is elaborately assessed and compared to the state of the art for a variety of test scenarios on data sets both semi-simulated from airborne HySpex data and real satellite-based WorldView-2 data.

Joint Sparse Representation-based Hyperspectral-Multispectral Image Fusion As was discussed in the Section 3.3, considerable progress and success has been achieved recently in the field of sparse representation-based hyperspectral and multispectral data fusion. This dissertation follows up on this development and further pushes the limits in what will be elaborately demonstrated to be one of the most reliable, quality-assuring approaches to enhancing the spatial resolution of hyperspectral data today – a demanding and very hot topic considering the upcoming hyperspectral satellite missions (see Tab. 1) and the vast possibilities brought by so far unattainable high-resolution hyperspectral satellite data.

The contribution of the second part of this dissertation is a new data fusion algorithm, called *Jointly Sparse Fusion of Hyperspectral and Multispectral Imagery* (J-SparseFI-HM), which combines the best elements of the existing approaches and adds new features that catapult the performance of the proposed algorithm to the top of the state of the art.

Key characteristics of J-SparseFI-HM relative to what exists today are summarized in the following.

- ◇ The fusion problem is mathematically formulated based on the sensor observation model introduced in [Eismann and Hardie, 2004], that is also underlying the majority of recently developed hyperspectral and multispectral fusion techniques [Yokoya et al., 2012; Wei et al., 2015a,c; Lanaras et al., 2015; Simões et al., 2015].
- ◇ One of the unique attributes of J-SparseFI-HM is its alternating three-stage local–non-local–global processing procedure: Patch-based reconstruction is implemented in one module combining local with non-local information retrieval, whereas a global processing module optimizes over the full spatial extent of the high-resolution input data, to

enhance spatial consistency across individually processed patches.

- ◇ Inspired by [Wei et al., 2015a] and [Simões et al., 2015], the global processing module of J-SparseFI-HM solves a regularized version of the inverse sensor observation model via convex optimization. Like in [Wei et al., 2015a], the corresponding penalty term incorporates the solution of the underlying sparse representation-based patch reconstruction process. The most elementary difference between the method proposed in [Wei et al., 2015a] and J-SparseFI-HM lies in that sparse representation-based patch reconstruction process as will be detailed below.
- ◇ In [Wei et al., 2015a], sparse reconstruction is conducted, feature band by feature band, on a spectral subspace, where the subspace dimension and spectral transformation matrix are computed *globally*. Consequently, less ordinary local features and rare materials are not retrievable in the fusion product as per definition. By contrast, J-SparseFI-HM is *highly adaptive* and uses *all* available system information and data including the full spectral dimensionality and, especially, the local spatial information to reconstruct high-resolution hyperspectral patches as accurately as possible.
- ◇ A sophisticated *Correlation-based HyperSpectral Grouping* (CorHySpeG) concept was developed to identify, for each patch-reconstruction task individually and automatically, an optimal set of partially overlapping groups of hyperspectral channels that are optimal for collective processing.
- ◇ The mutual correlation between hyperspectral channels in collectively processed groups is utilized as prior information for joint sparse recovery, which is elaborately demonstrated in Chapter 5 to improve the patch reconstruction accuracy compared to the individual processing of single channels.
- ◇ In [Wei et al., 2015a], for each PC-band, one *global*, *synthetic* dictionary is learned via ODL from a low-dimensional rough estimate of the target image. Those dictionaries are neither locally adaptive nor do they incorporate the nature of the scene. In J-SparseFI-HM, coupled dictionary atoms are sampled from the high-resolution input data (which is of higher resolution than any more or less rough estimate of the target image can possibly be) and a spatially degraded version of it for each reconstructed patch individually. Furthermore, for each of the locally varying groups of mutually correlated hyperspectral channels, one local dictionary is generated from a highly adaptive source image, which is computed, for that particular patch and group of channels, via non-negative linear regression of all multispectral channels.
- ◇ Local patch information is effectively utilized to split the patch reconstruction procedure into a local and a non-local processing part: First, for each channel group and patch, the optimum fit of the local patch in the adapted source data – represented as the first patch in the spatially and spectrally adapted dictionary – to the local input data is estimated via linear regression. This procedure produces a first hyperspectral patch estimate that fully utilizes not only local spatial patch information and the structural coherence between the input patches and produced patch, but also the full spectral information contained in both the multi- and hyperspectral input data. Subsequently, the residual between that local estimate and the original data is minimized via joint sparse recovery using the above-mentioned, appropriately sized, spatially and spectrally adaptive dictionaries that are composed of non-local, normalized, zero mean patches.
- ◇ In contrast to existing works, such as the methods proposed in [Grohnfeldt et al., 2014; Grohnfeldt and Zhu, 2015b; Wei et al., 2014, 2015a], much more system and data information is incorporated in the estimation procedure of the sparse coefficients: Instead of using only an estimate of the target image as measurement for sparse coefficient estimation, as done in [Wei et al., 2014, 2015a], both local input images are used as

measurements in addition to the estimation result of in the previous iteration. This is possible only by using coupled low- and high-dictionaries which allow accurate simultaneous fitting on both resolution levels to both input images. In comparison to all of the existing sparse representation-based hyperspectral and multispectral fusion methods, the sensor observation model and PSF- and SRF information incorporated therein are fully utilized in the estimation process of the sparse coefficients: In [Wei et al., 2014, 2015a], the sparse coding step is conducted solely on the *high-resolution level* using a synthetic dictionary and low-spectral-dimensional estimates of the target image as measurements; In [Grohnfeldt et al., 2013, 2014, 2015; Grohnfeldt and Zhu, 2015b], the (joint) sparse coefficient estimation process takes place solely on the *low-resolution level* using low-resolution dictionaries and the low-resolution hyperspectral input data as measurement. The proposed J-SparseFI-HM technique uses coupled dictionaries and, simultaneously, *both* the *high- and low-resolution* input data as well as the previous estimate of the full-spectral-dimensional high-resolution target patch as measurements – thus producing more reliable estimates of the (jointly) sparse coefficients than its predecessors.

- ◇ In contrast to the works in, e.g., [Grohnfeldt et al., 2013, 2014], the proposed J-SparseFI-HM algorithm is more versatile as it can be used for the processing of all types of hyperspectral and multispectral sensor combinations independently of their spectral overlap.

The proposed methodology is thoroughly validated in Sections 5.4 and 5.5. Detailed descriptions of adjustment parameters, comprehensive analyses of their impact on the quality of the data fusion product and the processing time, and recommendations for the default settings are provided. The quality of the data fusion products obtained via J-SparseFI-HM is elaborately assessed and compared to the state-of-the-art for a large variety of test scenarios using different sensor combinations, including a comprehensively generated pair of EnMAP and Sentinel-2 data.

4

The J-SparseFI Algorithm – A Solution to the Pan-Sharpening Problem

4.1 Introduction

This chapter presents the "Jointly Sparse Fusion of Images" (J-SparseFI) algorithm, which has been developed to solve the pan-sharpening problem. J-SparseFI is the first image fusion algorithm which exploits the inter-band mutual correlation via joint sparse recovery. Some of the contributions summarized in this chapter have been published recently in [Zhu, Grohnfeldt, and Bamler, 2016]. This chapter provides the latest update on the algorithm, new graphics as well as comprehensive descriptions, analyses and discussions.

The remainder of this chapter is organized as follows. Section 4.2 describes the data sets used to both illustrate the developed methodology and conduct experiments. Those data sets include WorldView-2-like images simulated from very high-resolution airborne visible/near-infrared (VNIR) HySpex hyperspectral data and real WorldView-2 data. The methodology part, including a modified version of the previously proposed SparseFI algorithm, a *joint sparsity model* (JSM) for the simultaneous processing of multiple bands, and the overall J-SparseFI algorithm, is described in Section 4.3. Section 4.4 discusses the influence of model-inherent system parameters and provides recommendations for reasonable settings. Section 4.5 contains a comprehensive comparative performance evaluation of J-SparseFI based on semi-simulated data. In Section 4.6, the practical use of J-SparseFI is demonstrated by pan-sharpening real WorldView-2 satellite imagery.

4.2 Data Sets

This section introduces the data sets used in this chapter. Two WorldView-2-like data sets, featuring different spatial resolution ratios, have been semi-simulated from HySpex [Lenhard et al., 2015b,a; Köhler, 2016] data using the simulation procedure introduced in Section 2.2.3. Those data sets are used in Sections 4.3 through 4.5 to assess and

analyze the performance of the proposed algorithm. Real WorldView-2 satellite data is used in Section 4.6 to further demonstrate the practical use of J-SparseFI.

4.2.1 WorldView-2 Images Simulated from HySpex Data

In order to accurately evaluate pan-sharpened imagery, it is common to work with simulated data. In this work, WorldView-2-like images are used, which have been simulated from airborne HySpex VNIR data acquired over Munich, Germany, in 2012 by experts at DLR. The HySpex data is of great value as, in addition to the high quality of the instrument itself, the acquired data is elaborately processed via a sophisticated processing chain at DLR. The HySpex VNIR sensor is characterized by a sub-meter GSD (if flown at a sufficiently low altitude) and 160 spectral channels ranging from 0.4 to 1.0 μm. The relative spectral responses of both HySpex and WorldView-2 are shown in Fig. 33. The 160 vertically oriented gray curves refer to the narrow spectral bands of the HySpex VNIR sensor. The bold colored curves indicate the SRFs of the satellite-based WorldView-2 multispectral instrument and the black curve shows the relative spectral response of the WorldView-2 panchromatic broadband sensor. It can be seen that the HySpex VNIR channels cover nearly the full spectral range of the WorldView-2 imager, which makes HySpex particularly suitable for the simulation of WorldView-2 and similar optical VNIR multispectral and panchromatic data.

The WorldView-2 data simulation has been conducted closely following the simulation procedure described in Section 2.2.3 and illustrated in Fig. 29 on page 45. In particular, the original high-resolution hyperspectral HySpex input data scene $\boldsymbol{I}_O \in \mathbb{R}^{N_b^O \times N_p^O}$ is characterized by $N_b^O = 160$ VNIR bands, $N_p^O = 3,600 \times 1,200 = 4{,}320{,}000$ pixels and a GSD of 0.75 m. The model degradation and error components have been constructed as follows.

- ◇ The spatial degradation matrix $\boldsymbol{S}^{O,Y} \in \mathbb{R}^{N_p^O \times N_p^Y}$ was constructed using the PSF approximation model detailed in Eq. (13) with a unit area A and a FWHM equalling the GSD of the WorldView-2 multispectral sensor relative to the GSD of the HySpex VNIR acquisition. Two data sets were simulated differing only in spatial resolution of the low-resolution multispectral data. The resolution ratios are set to 4 and 10. Hence, N_p^Y is equal to $N_p^O/4^2$ in one of the two simulated WorldView-2-like data sets and $N_p^O/10^2$ in the other one.
- ◇ For both data sets, the matrices $\boldsymbol{S}^{O,X}$ were equally set to the identity matrix $\mathbb{1}^{N_p^X \times N_p^X}$, where $N_p^X = N_p^O$.
- ◇ The spectral degradation matrices $\boldsymbol{R}^{O,X} \in \mathbb{R}^{1\times 160}$ and $\boldsymbol{R}^{O,Y} \in \mathbb{R}^{8\times 160}$ respectively contain the normalized WorldView-2 panchromatic and multispectral SRFs sampled to the center wavelengths of the HySpex channels.

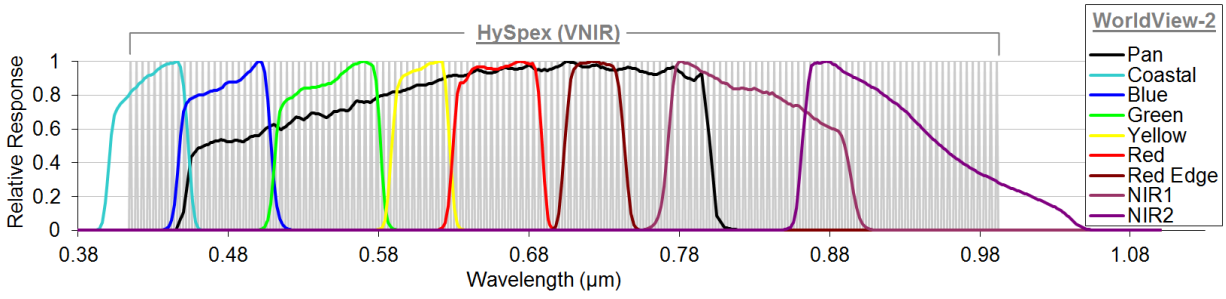
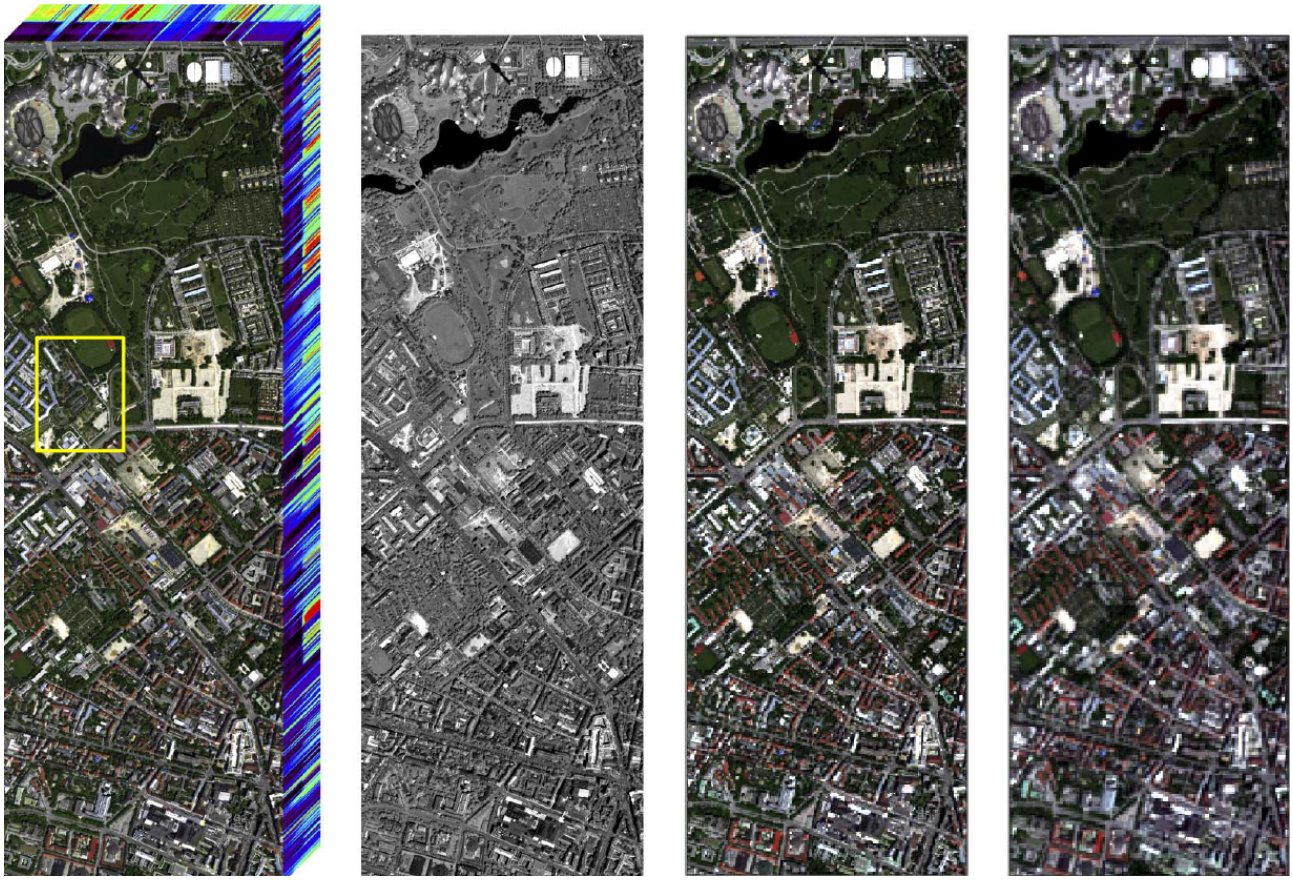


Fig. 33. Relative spectral responses of the HySpex VNIR sensor (gray narrow lines) and the WorldView-2 imager (colored curves indicate multispectral band SRFs and the black curve shows the SRF of the panchromatic sensor).



(a) I_O: HySpex VNIR Hyperspectral 0.75 m GSD (original)

(b) I_X: WorldView-2 Panchromatic 0.75 m GSD

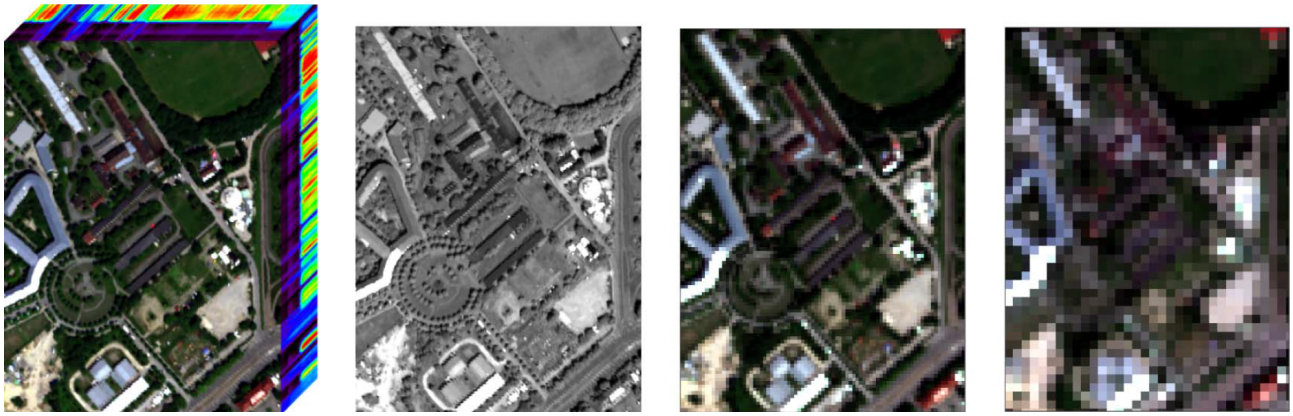
(c) I_Y: WorldView-2 Multispectral 3.0 m GSD

(d) I_Y: WorldView-2 Multispectral 7.5 m GSD

Fig. 34. Data sets derived from HySpex VNIR data. (a) Data cube of the original airborne 160-band HySpex VNIR data acquired over Munich, Germany, in 2012. (b) High-resolution panchromatic image simulated using the WorldView-2 spectral response. (c) Low-resolution multispectral image simulated using WorldView-2 spectral responses with a down-sampling factor of 4. (d) Low-resolution multispectral image simulated using WorldView-2 spectral responses with a down-sampling factor of 10. The bold yellow rectangle in (a) marks the area which is used for visual analysis later in this chapter.

- ◇ The additive error matrices $\boldsymbol{E}_X$ and $\boldsymbol{E}_Y$ were generated to simulate white Gaussian noise, so that $\boldsymbol{I}_X$ and $\boldsymbol{I}_Y$ feature SNR values between 10 db (strong noise) and "∞ db" (no added noise) as is extended on in Section 4.4.

The original HySpex hyperspectral data cube is shown in Fig. 34 (a). It has a spatial resolution of 0.75 m GSD and a spatial dimension of $3,600 \times 1,200$ pixels. Taking the hyperspectral image as input, synthetic low-resolution multispectral images and a high-resolution Pan image matching the specifications of the WorldView-2 imager in terms of the spectral properties are simulated. Fig. 34 (b) depicts the synthetic high-resolution panchromatic image, which has a spatial resolution equivalent to the hyperspectral data, i.e., 0.75 m GSD. As is shown in Tab. 1 on page 38, for most of the non-hyperspectral topographic satellites, the spatial resolution ratio F_{DS} between the high-resolution panchromatic and the lower-resolution multispectral sensor is equal to four. In order to be able to analyse data that is representative for those sensors, one of the two simulated data sets was generated with a resolution ratio of $F_{\mathrm{DS}} = 4$. Fig. 34 (c) shows the corresponding simulated low-resolution multispectral image of 3 m GSD. Moreover, in order to investigate the limits of the capabilities of J-SparseFI and other pan-sharpening methods, synthetic low-resolution multispectral data with a relatively high resolution ratio of $F_{\mathrm{DS}} = 10$ (i.e., 7.5 m GSD) has been simulated. This



(a) (b) (c) (d)

Fig. 35. Magnification of the area marked by the yellow rectangle in Fig. 34 (a). (a) Input hyperspectral data cube. (b) High-resolution panchromatic image. (c) Low-resolution multispectral image ($F_{DS} = 4$). (d) Low-resolution multispectral image ($F_{DS} = 10$). This test area will be used in Section 4.5 for the visual comparative analysis of the fusion results.

data is depicted in Fig. 34 (d). A reference high-resolution multispectral image with the original spatial resolution of 0.75 m GSD is generated using the spectral responses of the WorldView-2 imager. Comparing the pan-sharpening products to corresponding reference data allows for both quantitative and visual assessment of the fusion capacity. Moreover, reference information will be used to systematically analyse and evaluate the validity of the joint sparsity model.

Fig. 35 depicts a magnified view of the area marked by the yellow rectangle in Fig. 34 (a). This area was selected for visual analysis as it contains a diverse mixture of man-made objects and different types of vegetation. Figs. 35 (a) through (d) illustrate the input hyperspectral data cube, the simulated high-resolution panchromatic image, and the simulated lower-resolution multispectral images with down-sampling factors of 4 and 10, respectively. In Section 4.5, the visual analyses of the pan-sharpening results generated using J-SparseFI and other algorithms are performed using this test area, whereas statistical assessment is based on the full $3,600 \times 1,200$ pixel area. The unsigned 16-bit integer format of the original HySpex data is preserved in all data sets derived therefrom.

4.2.2 Real WorldView-2 Data

A real WorldView-2 image is used in Section 4.6 for practical demonstration of the proposed methodology. Fig. 36 illustrates a true-color RGB composition of that WorldView-2 8-band multispectral image, which was acquired over Munich on August 22, 2011. The scene contains $960 \times 1,320$ pixels with a spatial resolution of 2-m GSD. In order to have high-resolution reference data for comprehensive quality assessment of the reconstructed multispectral data, this image was used as original data to synthesize an 8-m low-resolution multispectral and a 2-m high-resolution panchromatic image as input imagery for the final pan-sharpening experiments. The sub-area marked by the yellow rectangle in Fig. 36 is selected for visual inspection, whereas the quantitative analysis is based on the full image information.

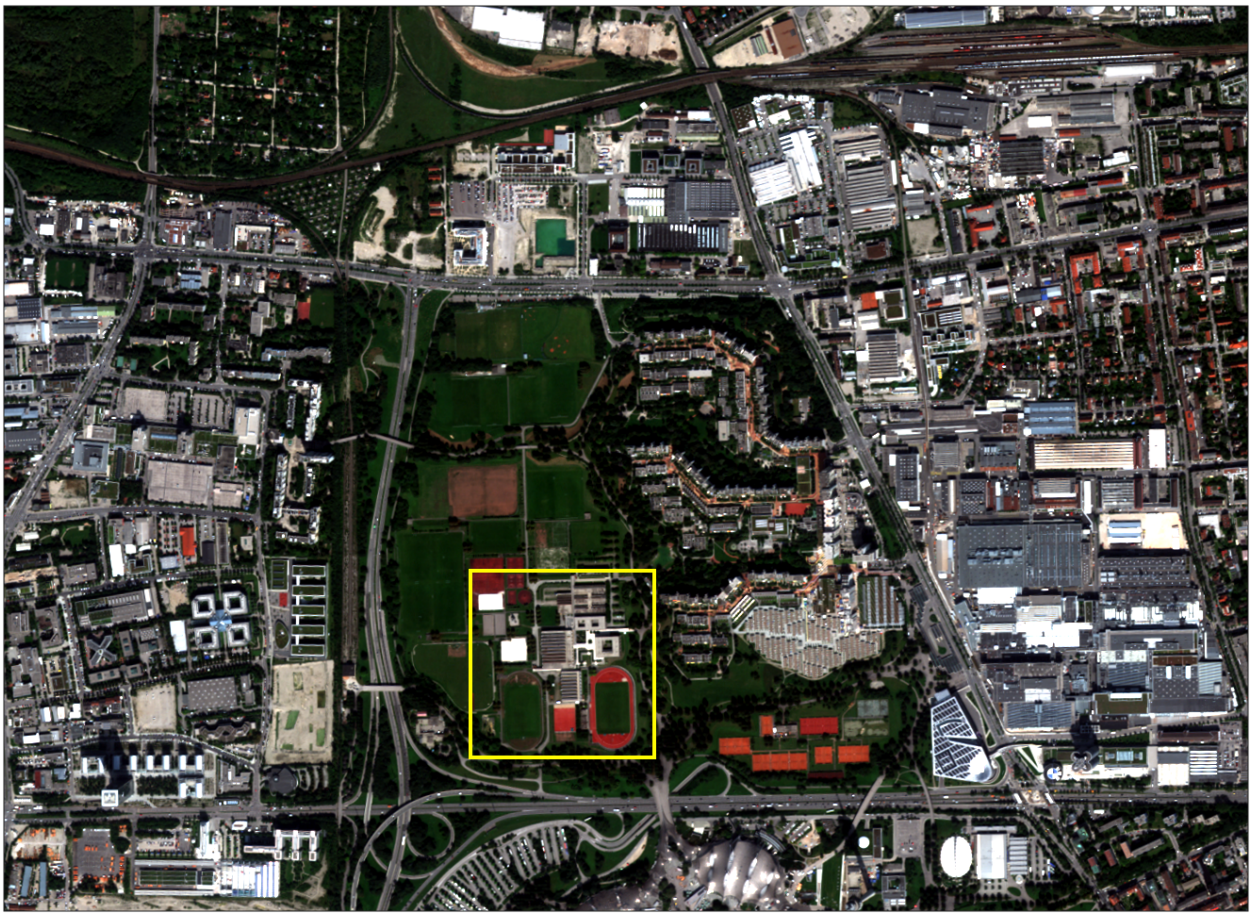


Fig. 36. True-color composition of a real WorldView-2 8-band multispectral image acquired over Munich on August 22, 2011. The scene contains 960 × 1,320 pixels at a GSD of 2 m.

4.3 Methodology

The J-SparseFI algorithm is based on a modified version of the sparse representation-based patch reconstruction approach implemented in SparseFI [Zhu and Bamler, 2013]. As one of the key elements of the proposed methodology, the *improved SparseFI algorithm* is briefly described in Section 4.3.1. Another key element is the proposed *joint sparsity model*, which is introduced in Section 4.3.2. With those two components at hand, Section 4.3.3 describes the J-SparseFI algorithm, which – in addition to the aforementioned modules – incorporates a joint sensor spectral response and channel mutual correlation analysis. This analysis is conducted prior to the actual pan-sharpening process. A representative, illustrative example for such analysis is detailed in Section 4.3.4.

4.3.1 Improved SparseFI Algorithm

The SparseFI algorithm was proposed by Zhu and Bamler [2013]. As an important element of the J-SparseFI algorithm, an improved version of the original SparseFI algorithm is proposed and introduced in this section. The main improvements lie in enhanced computational efficiency and practical usefulness for the processing of large data. First, in order to reduce the dimensionality of the individual patch optimization problems, the construction of the coupled dictionary pair has been changed as follows. In the original version of SparseFI, a *global* coupled dictionary pair is used, that is built up from *all* potential patches in the

image. This approach makes the dictionary size – and, thus, the dimension of the optimization problems – directly depending on the size of the input data. Hence, this concept not only is impractical for large-scale Earth observation data, but also makes systematic learning of the regularization parameter λ very challenging, whose optimum range depends on the size of the dictionary. One of the proposed improvements lies in changing from *global* to *local* coupled dictionary pairs. Those local dictionaries are composed of a predefined number of local coupled candidate patches as will be further explained below. As for the second improvement, the design of the original SparseFI algorithm requires a *sequential processing* of all patches in the image. This is because pixels that are located in overlapping parts of previously reconstructed high-resolution patches to the current patch under reconstruction, are integrated into the calculation of the latter. This inclusion of high-resolution pixels of neighboring, previously processed patches not only requires sequential computation, but also significantly increases the size of each dictionary atom, which, in turn, entails higher computational costs. Here, an efficient framework is proposed which reduces the size of the optimization problems and allows for a fully parallel processing of all patches without the necessity of cross-communication.

The pan-sharpening procedure starts with a low-resolution (LR) multispectral image $\boldsymbol{I}_Y$ with N_b^Y spectral channels and a co-registered high-resolution (HR) panchromatic (Pan) image $\boldsymbol{I}_X$. The goal is to increase the spatial resolution of $\boldsymbol{I}_Y$ while extending its spectral information to the sub-pixel grid given by the higher-resolution Pan image. That is, the fusion product is a HR multispectral image $\boldsymbol{I}_Z$ which is generated using both $\boldsymbol{I}_Y$ and $\boldsymbol{I}_X$ as inputs. The improved SparseFI algorithm consists of three main steps: 1) Dictionary learning, 2) sparse coefficient estimation and 3) HR multispectral image reconstruction.

Coupled Dictionary Learning The HR Pan image $\boldsymbol{I}_X$ is low-pass filtered and downsampled by a factor of F_{DS} (typically between 2 and 10) such that the resulting LR image $\boldsymbol{I}_{X,\mathrm{L}}$ features an identical spatial sampling grid and similar final point spread function to the multispectral channels. This step can be realized, for instance, by using a spatial degradation matrix $\boldsymbol{S}^{X,Y}$, which is generated following the procedure introduced in Section 2.2.3. The LR Pan image $\boldsymbol{I}_{X,\mathrm{L}}$, the HR Pan image $\boldsymbol{I}_X$ and the LR multispectral image $\boldsymbol{I}_Y$ are each tiled into N_p small, typically squared, possibly partially overlapping patches $\boldsymbol{x}_{\mathrm{L}}^p \in \mathbb{R}^{N_{\mathrm{L}}^P}$, $\boldsymbol{x}_{\mathrm{H}}^p \in \mathbb{R}^{N_{\mathrm{H}}^P}$ and $\boldsymbol{y}_k^p \in \mathbb{R}^{N_{\mathrm{L}}^P}$, where $k = 1,\ldots,N_b^Y$ indicates the k-th spectral channel, $p = 1,\ldots,N_p$ indicates the p-th image patch, and N_{L}^P and N_{H}^P denote the numbers of pixels in one low- and high-resolution patch, respectively. Note, that all originally 2-D spatial patches are reshaped to vector representations. For instance, a low-resolution Pan patch of size 5×5 corresponds to a vector $\boldsymbol{x}_{\mathrm{L}}^p$ containing $N_{\mathrm{L}}^P = 5^2 = 25$ elements.

A *local* low-resolution dictionary $\boldsymbol{D}_{\mathrm{L}}^p$, corresponding to the p-th patch undergoing reconstruction, is constructed from *some* subset of N_a, $N_a \leq N_p$, vectorized LR Pan patches extracted from $\boldsymbol{I}_{X,\mathrm{L}}$. Likewise, a high-resolution dictionary, $\boldsymbol{D}_{\mathrm{H}}^p$, is constructed from the corresponding N_a HR Pan patches extracted from $\boldsymbol{I}_X$. Before patches are added to the coupled high- and low-resolution dictionaries, $\boldsymbol{D}_{\mathrm{H}}^p$ and $\boldsymbol{D}_{\mathrm{L}}^p$, the mean value of the patch is subtracted and the patch is normalized by dividing each element by the L_2 norm of the HR patch. Hence, all dictionary atoms have a mean value of zero and the HR dictionary atoms all have an L_2 norm equal to 1. In this chapter, the local dictionary pair is composed of those N_a patches that are *spatially closest* to the target patch.

If $\{i_1^p,\ldots,i_{N_a}^p\} \subset \{1,\ldots,N_p\}$ denotes the index set of those N_a patches composing the local dictionaries which correspond to the p-th patch, those coupled dictionaries are generated as

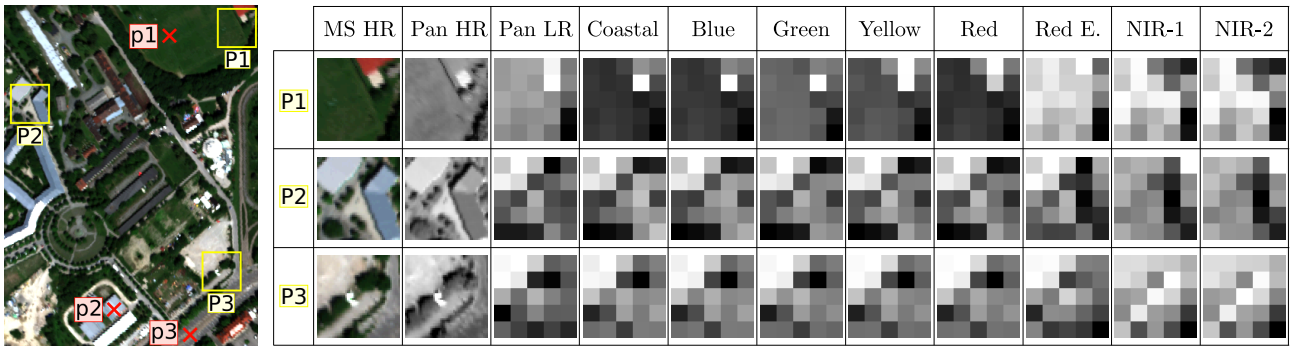


Fig. 37. Representative HR/LR dictionary atoms and corresponding LR multispectral image patches extracted from the data scene shown in Fig. 35. From top to bottom the patches show a primarily vegetated area, an urban area and a mixed area. From left to right, each line of patches corresponds to a true-color RGB representation of the HR multispectral image patch $\boldsymbol{Z}^p$, $\widetilde{\boldsymbol{x}}_{\mathrm{L}}^p$ from the LR dictionary $\boldsymbol{D}_{\mathrm{L}}^p$, $\widetilde{\boldsymbol{x}}_{\mathrm{H}}^p$ from the HR dictionary $\boldsymbol{D}_{\mathrm{H}}^p$, and $\boldsymbol{y}_k^p$, $k = 1,\ldots,8$ from the LR multispectral image. In this example, LR and HR patches are of sizes 5×5 and 50×50 pixels, respectively, which implies a GSD resolution ratio of $F_{\mathrm{DS}} = 10$.

follows:

$$\boldsymbol{D}_{\mathrm{L}}^p = \left[\, \widetilde{\boldsymbol{x}}_{\mathrm{L}}^{i_1^p}, \ldots, \widetilde{\boldsymbol{x}}_{\mathrm{L}}^{i_{N_a}^p} \,\right] \tag{84}$$

$$\boldsymbol{D}_{\mathrm{H}}^p = \left[\, \widetilde{\boldsymbol{x}}_{\mathrm{H}}^{i_1^p}, \ldots, \widetilde{\boldsymbol{x}}_{\mathrm{H}}^{i_{N_a}^p} \,\right] \tag{85}$$

where, for $i = i_1^p, \ldots, i_{N_a}^p$,

$$\widetilde{\boldsymbol{x}}_{\mathrm{H}}^i = \left(\boldsymbol{x}_{\mathrm{H}}^i - \overline{\boldsymbol{x}}_{\mathrm{H}}^i\right) \cdot \|\boldsymbol{x}_{\mathrm{H}}^i - \overline{\boldsymbol{x}}_{\mathrm{H}}^i\|_2^{-1} \tag{86}$$

and

$$\widetilde{\boldsymbol{x}}_{\mathrm{L}}^i = \left(\boldsymbol{x}_{\mathrm{L}}^i - \overline{\boldsymbol{x}}_{\mathrm{L}}^i\right) \cdot \|\boldsymbol{x}_{\mathrm{H}}^i - \overline{\boldsymbol{x}}_{\mathrm{H}}^i\|_2^{-1} \tag{87}$$

The reason for using the same divisor in Eqs. (86) and (87) is the essential correspondence between atoms in the coupled dictionary pair when multiplied by the *same coefficient* in the sparse reconstruction step described further below.

The current patch p undergoing reconstruction, should be always included in the corresponding local dictionary pair $(\boldsymbol{D}_{\mathrm{L}}^p, \boldsymbol{D}_{\mathrm{H}}^p)$, as this atom is likely to contribute the most to the reconstruction. Without loss of generality, the *first* atom in the local dictionary pair is assumed to correspond to the current patch under reconstruction, i.e. $i_1^p = p$.

Fig. 37 depicts representative atoms in the LR/HR dictionary pair and corresponding LR multispectral image patches, each of which is extracted from the data scene shown in Fig. 35. From top to bottom, the three patches show a primarily vegetated area, an urban area, and a mixed area. From left to right, the atoms $\widetilde{\boldsymbol{x}}_{\mathrm{L}}^p$ from the LR dictionary $\boldsymbol{D}_{\mathrm{L}}^p$, $\widetilde{\boldsymbol{x}}_{\mathrm{H}}^p$ from the HR dictionary $\boldsymbol{D}_{\mathrm{H}}^p$, and $\boldsymbol{y}_k^p$, $k = 1,\ldots,8$, from the LR multispectral image are illustrated. In this example, the LR and HR patches are of sizes 5×5 and 50×50 pixels, respectively. That is, the resolution ratio between the HR and LR GSDs is given by $F_{\mathrm{DS}} = 10$. By comparing the LR multispectral channels to the LR Pan image patch, it can be observed that, for each of the three areas, the Pan patch is highly correlated with the multispectral image patch in channels 1 (coastal) to 5 (red). This observation of mutual correlation will be utilized in a more general context by both the joint sparsity model introduced in Section 4.3.2 and in the correlation analysis described in Section 4.3.3. Moreover, the figure confirms that it is reasonable to represent the multispectral patches using the Pan patch, at least for these first five channels. Noticeable, channels NIR-1 and NIR-2 are not well represented by the Pan

image. Hence, SparseFI can be expected to suffer from representation issues at least in those two channels. This problem has been addressed and solved by the J-SparseFI algorithm as will be explained in Sections 4.3.2 and 4.3.4 and demonstrated in Sections 4.5 and 4.6.

Sparse Coefficient Estimation Let $\boldsymbol{y}_k^p$ be an LR multispectral patch, indexed as p, in the k-th channel. In order to produce its pan-sharpened high-resolution version $\boldsymbol{z}_k^p$, the above-constructed local HR and LR dictionary pair $(\boldsymbol{D}_\mathrm{L}^p, \boldsymbol{D}_\mathrm{H}^p)$ is employed as follows. Similarly to how the dictionary atoms are generated, the mean value $\overline{y}_k^p$ of the LR multispectral target patch $\boldsymbol{y}_k^p$ is subtracted, i.e.

$$\widetilde{\boldsymbol{y}}_k^p = \boldsymbol{y}_k^p - \overline{\boldsymbol{y}}_k^p, \tag{88}$$

where $\overline{\boldsymbol{y}}_k^p$ is the suitably-sized vector, whose coefficients are all equal to the mean value $\overline{y}_k^p$ of $\boldsymbol{y}_k^p$.

Next, $\widetilde{\boldsymbol{y}}_k^p$ is modeled as a linear combination of LR patches, i.e. atoms of the local dictionary $\boldsymbol{D}_\mathrm{L}^p$ weighted by a coefficient vector $\boldsymbol{\alpha}_k^p \in \mathbb{R}^{N_a}$. In order to calculate $\boldsymbol{\alpha}_k^p$ via sparse recovery, the number N_a of dictionary atoms has to be set such that $\boldsymbol{D}_\mathrm{L}^p$ is overcomplete. In this case, the linear system given by

$$\boldsymbol{D}_\mathrm{L}^p \boldsymbol{\alpha}_k^p = \widetilde{\boldsymbol{y}}_k^p \tag{89}$$

is underdetermined, which means that, under the assumption of $\boldsymbol{D}_\mathrm{L}^p$ having full row rank, there are infinitely many solutions to Eq. (89). The typical dimension of $\widetilde{\boldsymbol{y}}_k^p$ is between 9 (3×3) and 49 (7×7), and the typical number N_a of atoms in $\boldsymbol{D}_\mathrm{L}^p / \boldsymbol{D}_\mathrm{H}^p$ lies between 300 and 5,000. Since only far fewer atoms are needed to represent $\widetilde{\boldsymbol{y}}_k^p$, $\boldsymbol{\alpha}_k^p$ is very sparse. Accounting for this prior knowledge yields a significant dimensionality reduction of the solution space. For each LR multispectral patch $\widetilde{\boldsymbol{y}}_k^p$, the sparse coefficient vector $\boldsymbol{\alpha}_k^p$ is estimated via basis pursuit denoising (BPDN), which was introduced in Section 2.3.5. That is, l_1-regularized least squares is used to find a unique sparse coefficient vector:

$$\hat{\boldsymbol{\alpha}}_k^p = \arg\min_{\boldsymbol{\alpha}_k^p} \left\{ \lambda \|\boldsymbol{\alpha}_k^p\|_1 + \frac{1}{2} \left\| \boldsymbol{D}_\mathrm{L}^p \boldsymbol{\alpha}_k^p - \widetilde{\boldsymbol{y}}_k^p \right\|_2^2 \right\} \tag{90}$$

In order to overcome the amplitude bias introduced by BPDN, the estimate $\hat{\boldsymbol{\alpha}}_k^p$ can be updated using e.g. the SL1MMER algorithm developed by Zhu and Bamler [2010, 2012].

High-resolution Multispectral Image Reconstruction For the final generation of the HR multispectral image patches $\boldsymbol{z}_k^p$, the mean value $\overline{z}_k^p$ and the zero-mean variations,

$$\widetilde{\boldsymbol{z}}_k^p = \boldsymbol{z}_k^p - \overline{\boldsymbol{z}}_k^p, \tag{91}$$

are considered individually: The facts that $\boldsymbol{y}_k^p$ and $\boldsymbol{z}_k^p$ represent the exact same area in the same wavelength range, only at different spatial resolution scales, make it reasonable to assume that $\overline{y}_k^p = \overline{z}_k^p$. Furthermore, the coupled dictionaries are constructed and the pairs of HR/LR zero-mean atoms are jointly scaled in a way which suggests that $\widetilde{\boldsymbol{y}}_k^p$ and $\widetilde{\boldsymbol{z}}_k^p$ share the same sparse representation, by means of $\hat{\boldsymbol{\alpha}}_k^p$, in $\boldsymbol{D}_\mathrm{L}^p$ and $\boldsymbol{D}_\mathrm{H}^p$, respectively. Hence, the final pan-sharpened multispectral image patches $\boldsymbol{z}_k^p$ are reconstructed by replacing the low-resolution dictionary $\boldsymbol{D}_\mathrm{L}^p$ by the corresponding high-resolution one, $\boldsymbol{D}_\mathrm{H}^p$, in Eqs. (88)

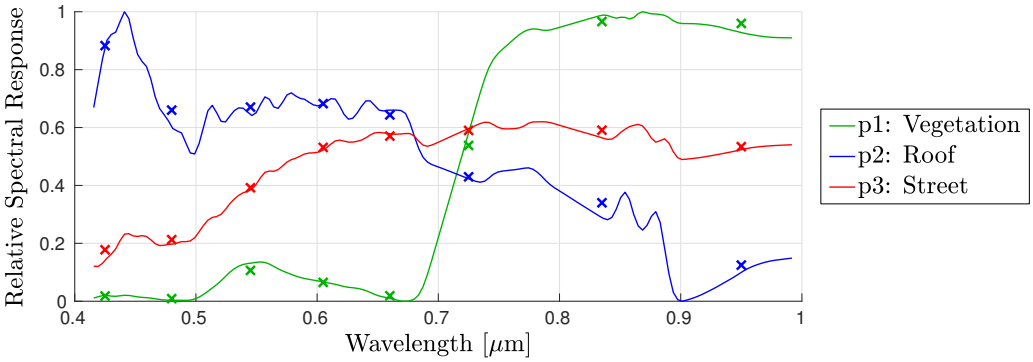


Fig. 38. Spectral profiles of three representative reflectance spectral profiles corresponding to different materials. The continuous curves indicate hyperspectral measurements and the discrete crosses indicate simulated WorldView-2 multispectral measurements.

and (89), yielding

$$\hat{\boldsymbol{z}}_k^p = \boldsymbol{D}_{\mathrm{H}}^p \hat{\boldsymbol{\alpha}}_k^p + \overline{\boldsymbol{y}}_k^p \,. \tag{92}$$

Tiling and averaging all reconstructed patches in the overlapping regions yields the final pan-sharpened image, $\boldsymbol{I}_Z$. Note, that – in contrast to the originally proposed SparseFI algorithm – the above-described framework allows a separate and, thus, fully parallel estimation of all individual sparse coefficient vectors $\hat{\boldsymbol{\alpha}}_k^p$, $p = 1,\ldots,N_p$.

4.3.2 Joint Sparsity Model (JSM)

Another important element of the J-SparseFI algorithm is the joint sparsity model (JSM). The JSM exploits the fact that one image patch acquired by different sensors or spectral bands may show spectral variations, but represents the same physical objects and, thus, structural patterns. Hence, it is plausible to assume that the same geometric features are present in all mutually correlated channels. For instance, in Fig. 37, one can observe that the channels number 1 (coastal) through 5 (red) are highly similar and that, in all three selected areas, the same geometrical structures are visible. The information obtained from this observation can be extended by analysing spectral profiles of different materials. In Fig. 38, the green, blue, and red curves indicate the spectral profiles, i.e., reflectance variations by wavelengths, of the pixels marked as p1 through p3 in Fig. 37. Those pixels correspond to nearly pure material spectra of a building roof (p1), green vegetation (p2), and a betuminized street (p3). The crosses along the colored curves in Fig. 38 mark the discrete values corresponding to the centers of eight multispectral WorldView-2 image bands. Noticeable, there is comparatively little variation between the leading five multispectral channels. This observation motivates the assumption that the coefficient vectors corresponding to the same patch in different channels are of similar structure.

Casting this basic idea into the sparse recovery framework, the theory suggests that the sparse coefficients $\hat{\boldsymbol{\alpha}}_k^p$ corresponding to correlated channels $\{k_1,\ldots,k_{N^Y_{b,,\mathrm{sub}}}\} \subset \{1,\ldots,N_b^Y\}$ are likely to be *jointly sparse*. That is, their supports, i.e. locations of their non-zero coefficients, are largely identical. However, the amplitudes of those non-trivial coefficients may vary arbitrarily from one channel to another.

Pan-sharpening offers an ideal basis for applying this theoretical model of joint sparsity. In this work, the evident high correlation between some of the multispectral channels is exploited by jointly estimating their sparse coefficient vectors using the joint sparse recovery tools introduced in Section 2.3. In particular, the modified SparseFI model is extended to

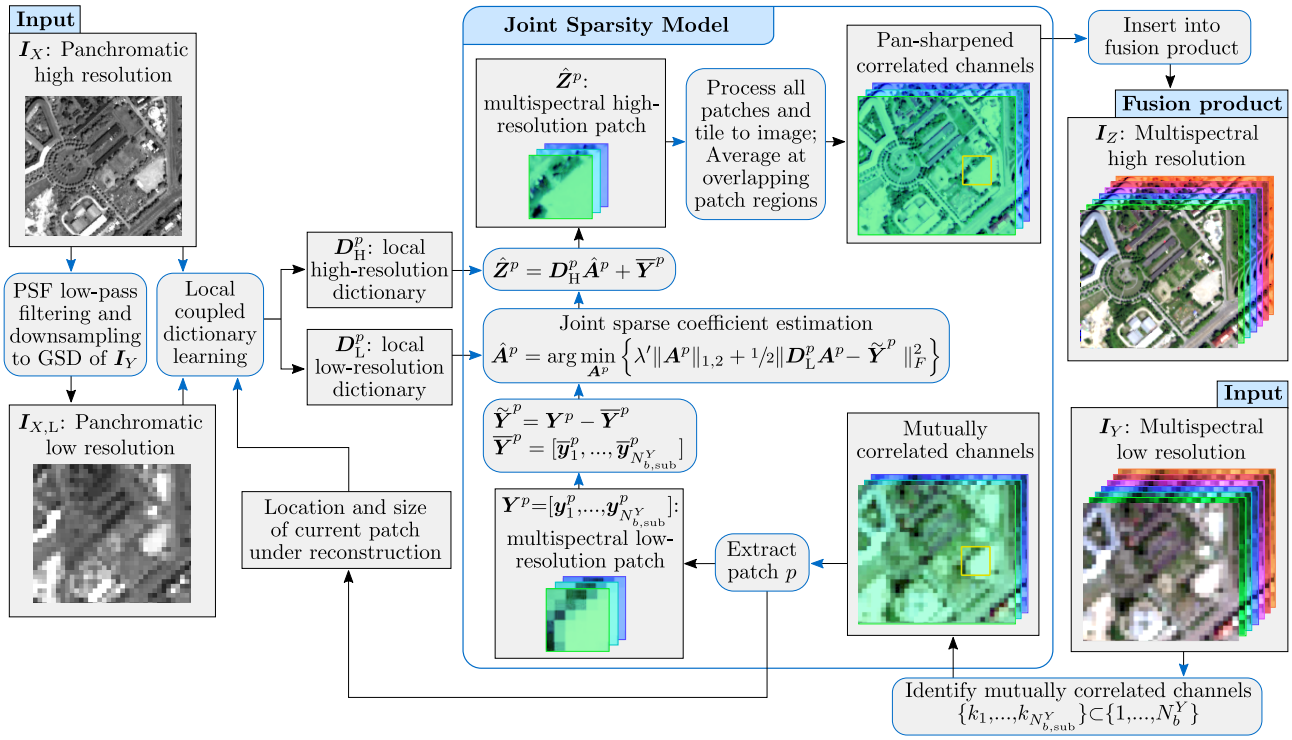


Fig. 39. Flowchart of the joint sparsity model (JSM), which is, here, depicted as a replacement of the sparse representation-based individual channel processing model embedded in the improved SparseFI algorithm, but can be used more generally as a stand-alone module, as is done by J-SparseFI.

the JSM as follows: The potentially high signal correlation between $N_{b,\mathrm{sub}}^Y$ ($N_{b,\mathrm{sub}}^Y \leq N_b^Y$) individual multispectral channels is utilized by the joint processing scheme of multiple channels shown in Fig. 39. In this flowchart, the JSM is depicted as a replacement of the sparse representation-based individual channel processing model embedded in the SparseFI algorithm. Although this flowchart shows the JSM as a core part of a SparseFI-like fusion scheme, it is considered more generally as a stand-alone module in J-SparseFI. The JSM requires as inputs (1) a high-resolution monochromatic image, (2) a group of mutually correlated bands extracted from a co-registered lower-resolution multispectral image, and (3) a coupled LR/HR dictionary pair. The core feature of the JSM lies in its sparse coefficient estimation step. In order to apply the theory of joint sparse recovery, the measurements, the sparse coefficients to be estimated, and the signals to be reconstructed are arranged as matrices as follows:

$$\tilde{\boldsymbol{Y}}^p = \left[\tilde{\boldsymbol{y}}_1^p, \ldots, \tilde{\boldsymbol{y}}_{N_{b,\mathrm{sub}}^Y}^p \right] \in \mathbb{R}^{N_L^P \times N_{b,\mathrm{sub}}^Y} \tag{93}$$

$$\hat{\boldsymbol{A}}^p = \left[\hat{\boldsymbol{\alpha}}_1^p, \ldots, \hat{\boldsymbol{\alpha}}_{N_{b,\mathrm{sub}}^Y}^p \right] \in \mathbb{R}^{N_a \times N_{b,\mathrm{sub}}^Y} \tag{94}$$

$$\tilde{\boldsymbol{Z}}^p = \left[\tilde{\boldsymbol{z}}_1^p, \ldots, \tilde{\boldsymbol{z}}_{N_{b,\mathrm{sub}}^Y}^p \right] \in \mathbb{R}^{N_H^P \times N_{b,\mathrm{sub}}^Y} \tag{95}$$

Note, that in Eqs. (93) through (95) the band indices $\{1, \ldots, N_{b,,\mathrm{sub}}^Y\} = \{k_1, \ldots, k_{N_{b,,\mathrm{sub}}^Y}\}$ are used without loss of generality for the sake of readability.

The joint estimation of the sparse coefficient vectors composing $\hat{\boldsymbol{A}}^p$ is done using the joint

sparsity extension of the BPDN optimization problem introduced in Eq. (74) on page 57:

$$\hat{\boldsymbol{A}}^p = \arg\min_{\boldsymbol{A}^p} \left\{ \lambda' \|\boldsymbol{A}^p\|_{1,2} + \frac{1}{2} \left\| \boldsymbol{D}_{\mathrm{L}}^p \boldsymbol{A}^p - \widetilde{\boldsymbol{Y}}^p \right\|_F^2 \right\} \tag{96}$$

The $l_{1,2}$-norm regularization term in Eq. (96) induces sparsity along the columns while minimizing the energy along the rows of a matrix. This cost function favors non-zero coefficients in the multispectral channels at the same positions. The following simple example illustrates this concept: Let $N_{b,\mathrm{sub}}^Y = 2$. Moreover, let $\hat{\boldsymbol{\alpha}}_1^p$ and $\hat{\boldsymbol{\alpha}}_2^p$ each contain a single non-zero coefficient, namely α_1 and α_2, respectively. If those two coefficients are located at different positions, then $\|\hat{\boldsymbol{A}}^p\|_{1,2} = |\alpha_1| + |\alpha_2|$. Otherwise the $l_{1,2}$-norm of $\hat{\boldsymbol{A}}^p$ is equal to $(\alpha_1^2 + \alpha_2^2)^{1/2}$, which is always smaller than the sum of the magnitudes. Hence, $l_{1,2}$-norm regularization favors row-sparse matrices, i.e. sparse column vectors with coefficients located at the same positions. As for the JSM, this observation means that image patches in different channels tend to be represented as linear combinations of the same atoms in the dictionary, but with different weights. The parameter λ' balances the joint sparsity level of the solution versus fidelity of the approximation $\boldsymbol{D}_{\mathrm{L}}^p \hat{\boldsymbol{A}}^p \approx \widetilde{\boldsymbol{Y}}^p$.

Similarly to how the single-channel image patches are reconstruction in SparseFI, the multichannel pan-sharpened image patches $\hat{\boldsymbol{Z}}^p$ are formed by the equation

$$\hat{\boldsymbol{Z}}^p = \boldsymbol{D}_{\mathrm{H}}^p \hat{\boldsymbol{A}}^p + \overline{\boldsymbol{Y}}^p, \tag{97}$$

in which $\overline{\boldsymbol{Y}}^p$ is defined by

$$\overline{\boldsymbol{Y}}^p = \left[\overline{\boldsymbol{y}}_1^p, \ldots, \overline{\boldsymbol{y}}_{N_{b,\mathrm{sub}}^Y}^p \right]. \tag{98}$$

Tiling all reconstructed patches $p = 1, \ldots, N_p$ to an image while averaging pixel values from adjacent overlapping patches yields the final image $\boldsymbol{I}_Z$, or, if $N_{b,\mathrm{sub}}^Y < N_b^Y$, a spectral subset of it.

In order for the JSM to be best effective, the high-resolution monochromatic image, from which the coupled dictionary pair is trained, should correspond to an SRF whose support widely intersects with the wavelength range covered by the SRFs of the jointly processed input multispectral bands. If so, the dictionary pair has the capacity of providing a physically meaningful representation of patches in all of the jointly processed multispectral channels. As was mentioned at the beginning of this chapter and illustrated in Fig. 33, this condition can not always be met. In the J-SparseFI algorithm described in the following section, such physical aspects and practical considerations are taken into account.

4.3.3 The J-SparseFI Algorithm

As was mentioned in Section 3.3, methods that directly train the dictionary pair from the input Pan image, such as SparseFI and its two-step extension, can be applied to a broad range of image fusion problems. However, they require the Pan and multispectral images to cover the same wavelength region, which is a condition not met by most of the modern satellite-based sensors (see Tab. 1 on page 38). The J-SparseFI algorithm described below incorporates a sophisticated fusion strategy, which copes with the challenging scenario of the Pan and the multispectral bands not being responsive in the same wavelength regions or not even being statistically correlated.

Moreover, the JSM implemented in J-SparseFI is designed to exploit the mutual correlation

Algorithm 3 Sensor spectral response and channel mutual correlation analysis

Input: Multispectral and Pan sensor SRFs, low-resolution multispectral image $\boldsymbol{I}_Y$, PSF-filtered and down-sampled Pan image $\boldsymbol{I}_{X,\mathrm{L}}$, correlation threshold θ

Calculate the channel mutual correlation matrix $\boldsymbol{\mathcal{C}} := \boldsymbol{\mathcal{C}}([\boldsymbol{I}_Y, \boldsymbol{I}_{X,\mathrm{L}}])$
Identify θ-*blocks*[†] in $\boldsymbol{\mathcal{C}}$
for each θ-block **do**
    for each (multispectral) channel k in the block **do**
        if multispectral SRF in channel k is **not** *well covered*[‡] by the Pan SRF **then**
            Exclude channel k from the block;
        end if
    end for
    if updated block size $\geq 2 \times 2$ **then**
        Identify current block as *primary group* of channels for which the dictionaries are trained from $(\boldsymbol{I}_X, \boldsymbol{I}_{X,\mathrm{L}})$;
    end if
end for
Exclude channels corresponding to primary groups from list of channels to be assigned and reduce $\boldsymbol{\mathcal{C}}$ to $\boldsymbol{\mathcal{C}}'$ accordingly;
Identify θ-blocks in $\boldsymbol{\mathcal{C}}'$ and assign each block as *secondary group*;
Assign all remaining multispectral channels as *individual channels*;

Output: Division of all multispectral channels into *primary groups*, *secondary groups* and *individual channels*.

[†] Define θ-*block* – or simply *block* – in a correlation matrix $\boldsymbol{\mathcal{C}}$ as a square sub-matrix of size 2×2 or larger, centered along the diagonal of $\boldsymbol{\mathcal{C}}$, whose elements are all greater than or equal to θ; A block corresponds to a group of adjacent spectral channels.

[‡] Suppose that the integrals of the SRFs of both the Pan and the multispectral channel under consideration are equal to one. Then, the multispectral channel SRF is considered *well-covered* by the Pan SRF if the integral of the product of both SRFs is greater than 0.05.

between adjacent channels for fusion quality enhancement via joint sparse signal recovery.

The workflow of J-SparseFI is shown in Fig. 40. As can be seen, it jointly uses the modified SparseFI algorithm and the proposed JSM as modules to process different groups of channels. The overall procedure incorporates four main steps: (1) Sensor spectral response and channel mutual correlation analysis, (2) the JSM to jointly process channels assigned to *primary groups*, (3) the modified SparseFI to process *individual channels*, and (4) the JSM to jointly sharpen channels assigned to *secondary groups* of channels. Each of these key components is briefly described in the following.

Sensor Spectral Response and Channel Mutual Correlation Analysis In addition to the above-described modified SparseFI algorithm and the JSM model, one key element of the J-SparseFI algorithm is a sensor spectral response and channel mutual correlation analysis. The goal of this analysis is to divide all N_b^Y multispectral channels into groups each of which belongs to one of the following categories:

- ◇ *Primary group* of channels to be jointly processed: Group of adjacent channels with high mutual correlation whose SRFs are responsive in wavelength ranges well covered by the Pan sensor SRF;
- ◇ *Secondary group* of channels to be jointly processed: Group of adjacent channels with high mutual correlation which correspond to wavelength ranges outside or partially outside of the support of the Pan sensor SRF;
- ◇ *Individual channels*: Channels assigned to neither primary nor secondary groups. Those channels are individually processed using the improved SparseFI algorithm.

The detailed procedure of this analysis is described in Algorithm 3.

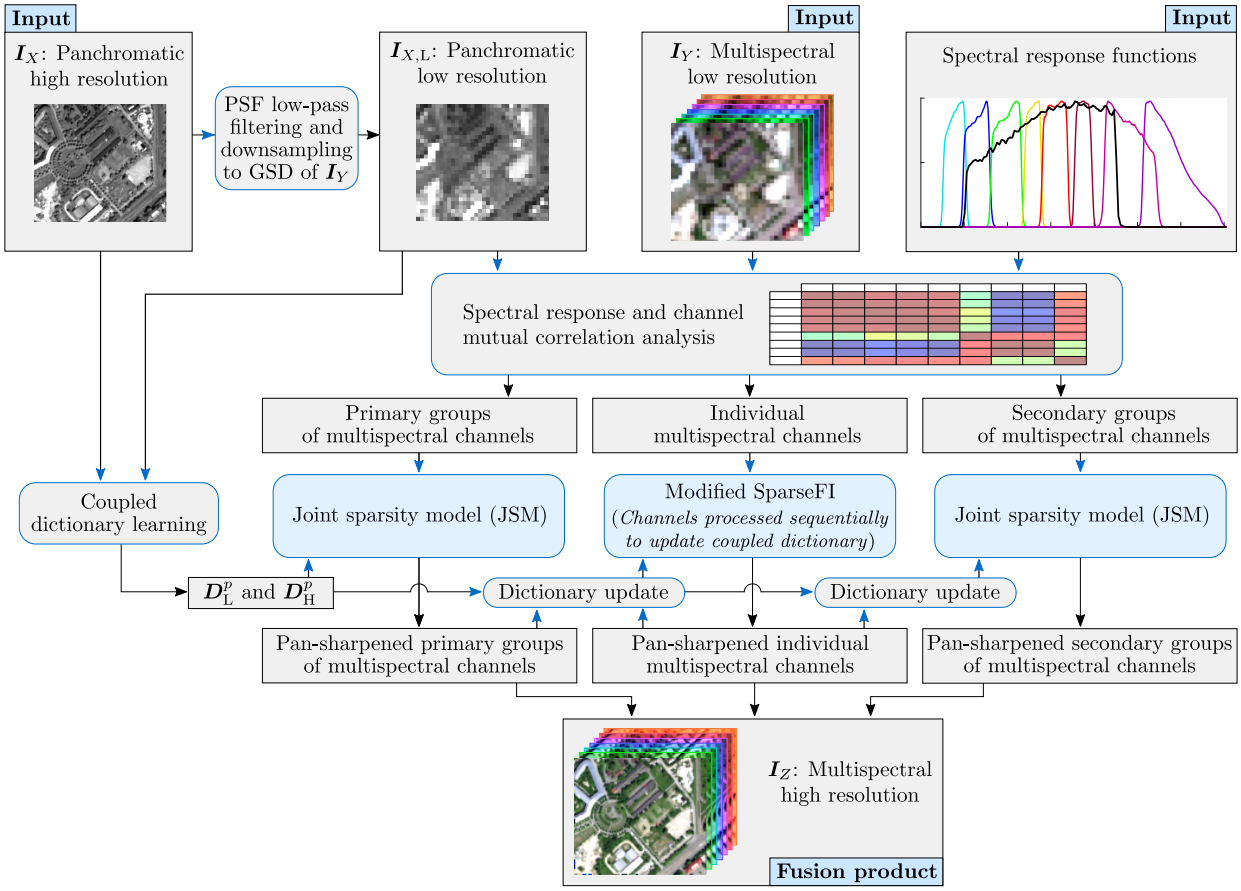


Fig. 40. Flowchart of the J-SparseFI algorithm. A step-by-step description of the procedure depicted in this figure is given in Section 4.3.3.

Processing Primary Groups of Channels via JSM To overcome the problem caused by the common mismatch between the wavelength ranges of the Pan and multispectral channels, J-SparseFI first processes the mutually correlated channels well covered by the Pan SRF using the Pan image for dictionary training and the JSM for patch-based reconstruction.

Subsequently, the pan-sharpened results of those primary groups are potentially used instead of the original Pan image as input for the dictionary training step, when processing those channels less well represented by the Pan image. The decision about which of the sharpened multispectral channels or Pan image is used for dictionary training is automatically derived from the channel mutual correlation matrix $\mathcal{C}([\boldsymbol{I}_Y, \boldsymbol{I}_{X,L}])$ (defined in Section 2.1.6 on page 31) calculated from all low-resolution multispectral channels and the PSF-filtered and down-sampled Pan image. An example of a mutual correlation matrix in the case of WorldView-2 data is given in Fig. 41, which will be further discussed in Section 4.3.4.

Processing Individual Channels via the Improved SparseFI Module After the primary groups are pan-sharpened, the individual channels are processed next, one channel after another. The rationale behind the sequential processing is that it allows to continuously update the high-resolution source data used for dictionary training and, therefore, improve the quality of the overall fusion product.

Processing Secondary Groups of Channels via JSM Similarly to how the individual channels are processed sequentially, the secondary groups are processed at last, one after another while updating the source data used for dictionary training. In contrast to individual channels, however, the channels contained in each secondary group are highly mutually corre-

	first θ-block: individual channel	first θ-block: primary group				second θ-block: individual channel	second θ-block: secondary group		
	Coastal	Blue	Green	Yellow	Red	Red Edge	NIR-1	NIR-2	**Pan**
Coastal	1	0.995	0.952	0.956	0.947	0.419	-0.075	-0.074	0.811
Blue	0.995	1	0.971	0.969	0.955	0.443	-0.051	-0.050	0.832
Green	0.952	0.971	1	0.973	0.942	0.559	0.085	0.087	0.898
Yellow	0.956	0.969	0.973	1	0.992	0.510	-0.002	-0.001	0.875
Red	0.947	0.955	0.942	0.992	1	0.467	-0.056	-0.055	0.842
Red Edge	0.419	0.443	0.559	0.510	0.467	1	0.853	0.852	0.860
NIR-1	-0.075	-0.051	0.085	-0.002	-0.056	0.853	1	0.999	0.480
NIR-2	-0.074	-0.050	0.087	-0.001	-0.055	0.852	0.999	1	0.481
Pan	0.811	0.832	0.898	0.875	0.842	0.860	0.480	0.481	1

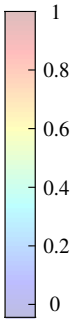


Fig. 41. Channel mutual correlation matrix $\boldsymbol{\mathcal{C}}([\boldsymbol{I}_Y \boldsymbol{I}_{X,\mathrm{L}}])$, calculated from the synthesized WorldView-2 data introduced in Section 4.2.1. The indicated θ-blocks, primary group of channels, secondary group of channels and individual channels have been automatically identified following the procedure described in Algorithm 3.

lated and, therefore, processed jointly using the JSM. Since the channels in the secondary groups are known to be less well represented by the original Pan image, it is crucial to create the dictionary from the previously produced high-resolution multispectral band data which best approximates the average band within the secondary group under consideration. In the experiments presented in Sections 4.5 and 4.6, this step will demonstrate to significantly improve the quality of the fusion results in the secondary groups.

4.3.4 J-SparseFI Applied to WorldView-2-Like Data

The proposed J-SparseFI fusion scheme can be applied to data acquired by all kinds of multiresolution multispectral/Pan sensor pairs. Currently operational potential satellite-based instruments include, but are not limited to, Landsat-8, WorldView-2 through 4, Spot-7, Pléiades, and GeoEye-1. A more comprehensive list of sensors and their characteristics can be found in Tab. 1 on page 38. In this section, the above-introduced analysis and individual processing steps of J-SparseFI are exemplified using the WorldView-2-like data described in Section 4.2.1.

Fig. 41 illustrates the channel mutual correlation matrix $\boldsymbol{\mathcal{C}}([\boldsymbol{I}_Y, \boldsymbol{I}_{X,\mathrm{L}}])$ calculated from the 8-band low-resolution multispectral and the PSF-filtered and down-sampled Pan input data. Since the difference in the mutual correlation between different channels under the resolution ratios $F_{\mathrm{DS}} = 4$ and 10 is negligible, input images with a resolution ratio of 10 are selected in this table without loss of generality.

Following the procedure described in the previous section, the sensor spectral response and channel mutual correlation analysis is conducted based on the relative spectral responses of the WorldView-2 imager (see Fig. 33) and the channel mutual correlation matrix $\boldsymbol{\mathcal{C}}([\boldsymbol{I}_Y, \boldsymbol{I}_{X,\mathrm{L}}])$ shown in Fig. 41. According to Algorithm 3, line 2, the channels number 1 through 5 and 7 through 8 are identified as the only two θ-*blocks*, where θ is set to 0.9. That is, each group has an overall mutual correlation greater than or equal to 0.9.

Continuing with lines 3 through 12, in the first θ-block, only channel number 1 (coastal) is not well covered by the Pan SRF and is, thus, excluded from its block (see lines 5 and 6). Hence, channels number 2 through 5 (blue, green, yellow, and red) are identified as the primary group of channels. Channels number 7 and 8 (NIR-1 and NIR-2), which compose the second θ-block, are both not well covered by the Pan SRF. Hence, channels number 2 through 5 remain the only primary group. After excluding this group from the list of

channels to be assigned, channels number 7 and 8 (NIR-1 and NIR-2) are identified as the only secondary group (see lines 13-14). Consequently, channels number 1 (coastal) and 6 (red edge) remain to be identified as individual channels (see line 15).

Now, that the initial analysis is done, the primary group, individual channels, and secondary group of channels are processed in a sequential manner. First, the primary group of channels is sharpened via JSM using local coupled dictionary pairs built up from the original high- and low-resolution Pan image pair. This step yields high-resolution versions of the multispectral channels blue, green, yellow, and red. Next, the coastal channel is individually processed via the modified SparseFI algorithm using an updated coupled dictionary pair. As a source high-resolution image for the dictionary learning, the previously reconstructed blue channel and its PSF-filtered and down-sampled version are used instead of the original Pan image. This is because, among the Pan and the channels composing the primary group, the blue channel correlates the most with the coastal one (see Fig. 41). Similarly, the red-edge channel is individually processed via the modified SparseFI algorithm using a dictionary pair trained from the original low- and high resolution Pan image pair, because there is no multispectral channel which has a higher correlation to the red-edge channel than the Pan image (see Fig 41. Finally, the NIR-1 and NIR-2 channels composing the secondary group are jointly reconstructed via the JSM using a dictionary pair built of from the previously reconstructed high-resolution red-edge channel and its PSF-filtered and down-sampled version.

4.4 Recipe for Choosing the Tuning Parameters

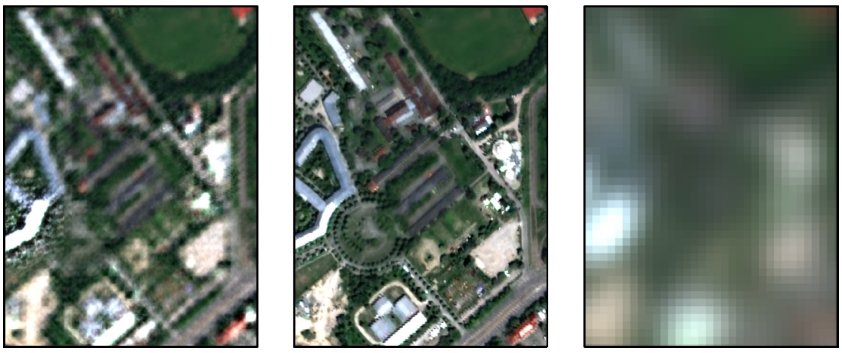
Given the availability of reference data in the case of the WorldView-2 imagery synthesized from the HySpex data, it is possible to systematically analyze the impact of the tuning parameters involved in the proposed J-SparseFI algorithm. This section contains such analyses based upon which a recipe describing how to best select parameters is provided. The parameters discussed in the following subsections include the regularization parameters λ and λ', the number N_{L}^{PO} of overlapping pixels between adjacent patches (specified on the low-resolution scale), and the dictionary size N_a given by the number of dictionary atoms/patches.

4.4.1 Means of Fusion Quality Evaluation

In order to quantitatively assess the quality of the fusion products generated by J-SparseFI and comparative pan-sharpening methods in the scope of the experiments conducted in this chapter, the fusion products are compared to reference data using established assessment metrics. Specifically, the following metrics have been selected for quality evaluation in this chapter, which are all defined in more detail in Section 2.2.4:

- ◇ Correlation coefficient (CC)
- ◇ Relative dimensionless global error in synthesis (ERGAS)
- ◇ Spectral Angle Mapper (SAM)

The reason for selecting these three metrics is that, in combination, they balance an assessment of spectral and spatial information preservation. Since the SAM measures the spectral angle between normalized spectral profiles accumulated over all pixels, it primarily indicates *spectral distortion* of the fusion product relative to the reference data. In contrast, the CC measures similarity between two multispectral images *band by band*. It is, therefore,



(a) (b) (c)

Fig. 42. True-color RGB compositions of subimages extracted from high-resolution multispectral fusion results, produced using (a) $\lambda' = 0$, (b) $\lambda' = 1$ and (c) $\lambda' = 10^6$. The underlying input data is the synthesized WorldView-2 data with resolution ratio of $F_{\mathrm{DS}} = 10$. Other system parameters used for processing are as follows: Dictionary size $N_a = 600$, patch size $N_{\mathrm{L}}^{PS} = 5^2$ low-resolution pixels, and patch overlap $N_{\mathrm{L}}^{PO} = 4$ low-resolution pixels.

considered a measure for *spatial distortion*. The average CC over all channels indicates the overall spatial discrepancy between the produced and the "ideal" images. Lastly, the ERGAS metric is originally defined as a quality measure for multi-band data (see page 46). It is based on the root mean square error (RMSE) metric and accounts for the resolution ratio F_{DS}. The net (total) ERGAS value jointly indicates spectral and spatial errors. In addition to the original net ERGAS metric, which is applied to the full multi-band image, the ERGAS measure can be applied to all bands separately. This allows for a more in-depth analysis of the fusion performance in each band and is, therefore, well suited for evaluation of the proposed J-SparseFI fusion concept.

4.4.2 Regularization Parameters

The regularization parameters λ (used in the modified SparseFI module) and λ' (used in the JSM) are key tuning parameters in J-SparseFI. Since λ and λ' affect the reconstruction results in a similar way, this section is limited to a discussion of λ'. This regularization parameter balances data fidelity versus sparsity level in the optimization problem stated in Eq. (96). As is depicted in Fig. 42 (a), in the case of $\lambda' = 0$, Eq. (42) degenerates to the least squares solution, whereas it converges to the zero solution as $\lambda' \to \infty$. Fig. 42,(c), illustrates the latter case in which $\hat{\boldsymbol{A}}^p \to 0$ for all $p = 1, \ldots, N_p$. Hence, it follows from Eq. (97), that all high-resolution pixels in each multi-channel patch $\hat{\boldsymbol{Z}}^p$ equally take the mean values stored in $\overline{\boldsymbol{Y}}^p$.

The choice of λ' significantly influences both the fusion quality and the computational cost. Those two aspects are illustrated in Figs. 43 (a)-(c) and Fig. 44 (a), respectively. In [Strait et al., 2008], it is suggested to select λ' according to the noise standard deviation σ and the dictionary size N_a. As will be addressed later in this section, the favorable range for N_a is rather limited. The impact of the noise level on the dependency of the fusion quality on λ' is depicted in Fig. 43. The plots show – from left to right – the fusion quality assessed using (a) the average CC, (b) total ERGAS, and (c) SAM, as functions of λ' under different SNR scenarios. In this experiment, the dictionary size, i.e., the number of atoms, is fixed to $N_a = 600$. Tests have been conducted with λ' ranging from 10^{-10} to 10^5. The presented plots allow for the following conclusions:

- ◇ Nearly independently of the SNR, there is a broad peak (CC) and trough (ERGAS) parameter space ranging from 10^{-2} to 10^2. This observation reveals that, within this wide

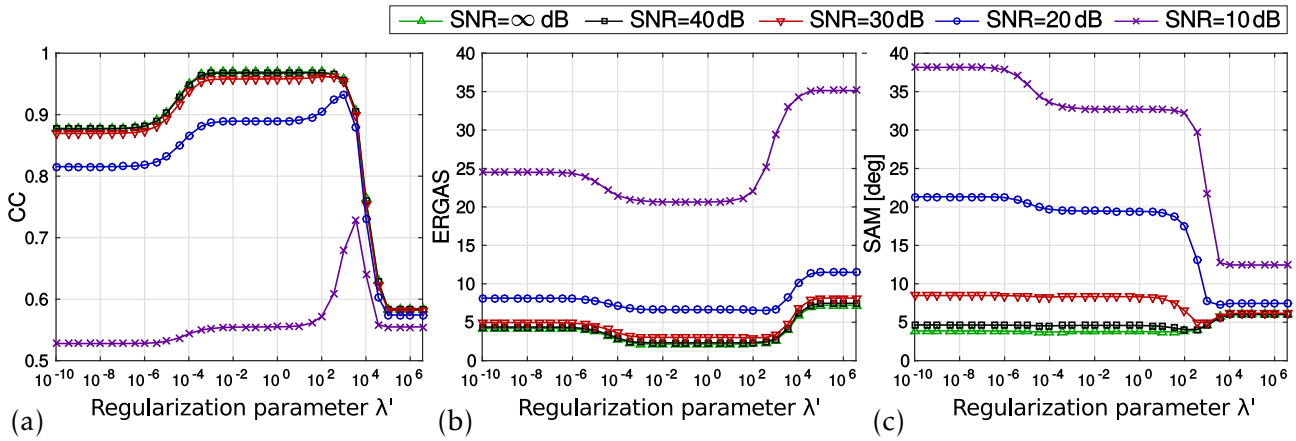


Fig. 43. Analysis of the impact of changes in the joint sparsity regularization parameter λ' and the SNR of the input data on the overall fusion quality, which is assessed using (a) CC, (b) ERGAS and (c) SAM as quantitative assessment metrics. Input data and system settings are as described in the caption of Fig. 42.

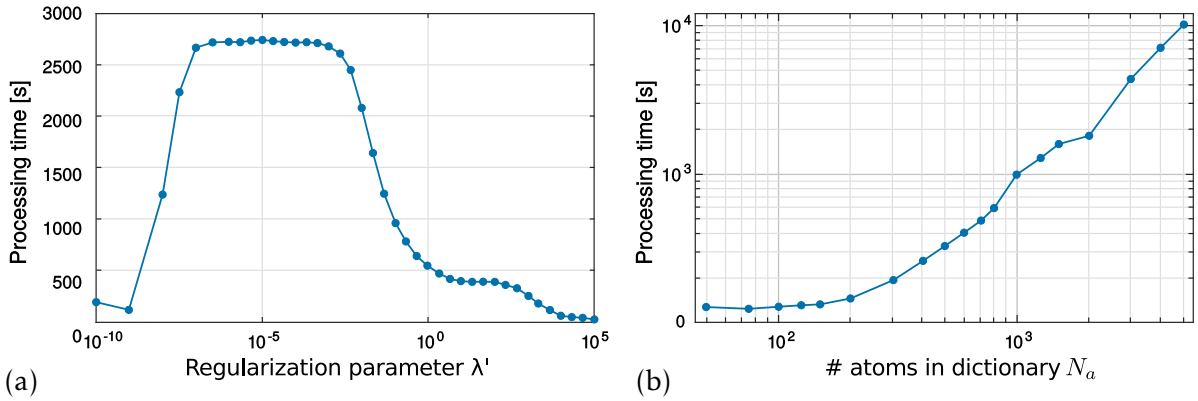


Fig. 44. Overall processing time measured under varying (a) regularization parameter λ' and (b) dictionary size N_a. The patch coefficients were reconstructed in parallel using 128 cores. Input data and system settings are all as described in the caption of Fig. 42.

region of moderate λ' values, a solid, stable and good performance can be expected. As $\lambda' \to 0$, the sparsity level of the coefficient vectors decreases and the fusion performance worsens as $\hat{\boldsymbol{A}}^p$ approaches the (dense) least squares solution. On the other hand, for $10^3 < \lambda \to \infty$ the solution $\hat{\boldsymbol{Z}}^p$ increasingly tends to $\overline{\boldsymbol{Y}}^p$, which yields even lower fusion quality than the least squares solutions. This analysis evidences the performance improvements achievable through regularizing the cost function in Eq. (96) by the joint sparsity promoting $l_{1,2}$-term.

- ◇ There is no optimum value for λ' which would simultaneously lead to best fusion quality assessment numbers in all three metrics. For example, the optimum λ' value w.r.t. CC is not the optimum value w.r.t. ERGAS and SAM, and vice versa.
- ◇ Noticeably, although the optimum λ' values for the different quality metrics are different and generally vary for different SNRs, as long as $\lambda \in [10^{-2}, 10^2]$, J-SparseFI yields stable and good results w.r.t. all metrics except for SAM at high SNRs.

Based on the preceding analysis, a moderate value of λ' between 10^{-2} and 10^2 is recommended if the primary goal is to produce pan-sharpening results of good *quality*.

In addition to image quality, another important aspect is the computational cost. Fig. 44 (a) indicates the computational time (measured using 128 cores) as a function of λ'. Within the aforementioned moderate parameter range, the computational time drops significantly as λ' increases. A similar observation can be made as λ' lowers from 10^{-7} to 10^{-10}. This impact of λ' on the processing time is a property of the utilized *Fast Iterative Soft Thresholding*

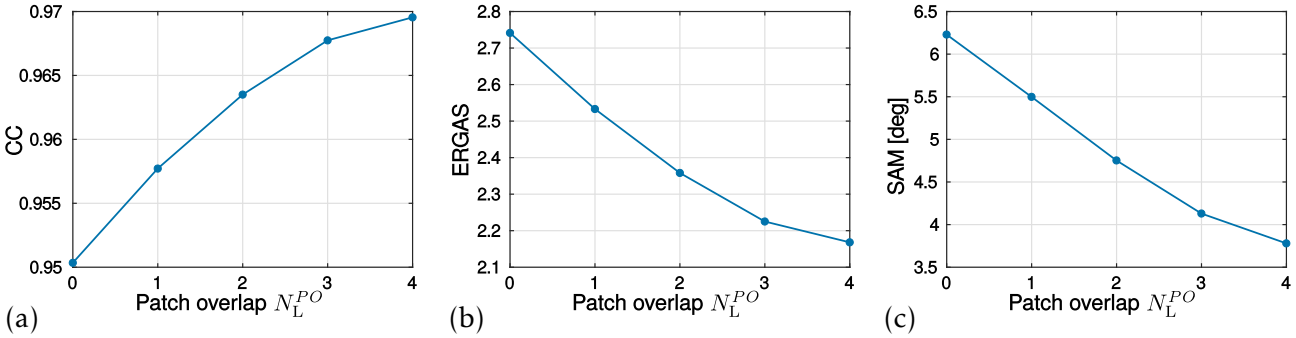


Fig. 45. Analysis of the impact of changes in the patch overlap N_L^{PO} on the overall fusion fidelity, which is assessed using (a) CC, (b) ERGAS and (c) SAM as quantitative assessment metrics. Except for N_L^{PO}, all system settings are as described in the caption of Fig. 42.

Algorithm (FISTA) described in Section 2.3.7, which typically requrires fewer iterations both if the solution is very sparse ($\lambda' \to \infty$) and if the solution converges to the simple least squares solution ($\lambda \to 0$).

Finally, the optimum λ' value, particularly w.r.t. CC and computational time, can be found on the upper end of the interval $[10^{-2}, 10^{2}]$. However, beyond this potentially small subrange, the fusion quality drops rapidly. Moreover, the exact optimum parameter ranges may slight vary from one data set to another, even though not significantly, since all patches are normalized and have zero-mean. Hence, it is recommended to choose λ' conservatively between 10^0 and 10^1.

4.4.3 Patch Overlap

For sparse representation-based pan-sharpening methods such as the one proposed in [Zhu and Bamler, 2013], which process image patches in a sequential manner and require previously reconstructed overlapping high-resolution image patches as inputs for subsequent patch reconstructions, the overlapping area is recommended to be set no larger than 20% to 40% of the patch size. In contrast, J-SparseFI processes each patch independently. This approach not only allows for full parallelization, but also simplifies each optimization problem. In the final step, all reconstructed patches are averaged in the overlapping areas. Here, the impact of the relative patch overlap N_L^{PO} on the overall quality of the fusion result is analyzed.

In this experiment, the patch size is set to $N_L^P = 5^2$ (i.e. 5×5) low-resolution pixels. The regularization parameter is set to $\lambda' = 1.0$ (center of the stable region). Furthermore, the dictionary size is set to $N_a = 600$ atoms and the resolution ratio of the input images is selected as $F_{DS} = 10$.

Varying patch overlapping of $N_L^{PO} = 0$ through 4 low-resolution pixels is tested, where N_L^{PO} equally corresponds to both spatial directions. In particular, that $N_L^{PO} = 0$ means no patch overlap and $N_L^{PO} = 4$ is the maximum possible patch overlap. Similarly to the testing discussed in the previous subsection, Fig. 45 illustrates the fusion quality assessment as a function of N_L^{PO}. As can be expected, the performance monotonically improves with increasing patch overlap. This exact observation has been made consistently in various unpublished experiments conducted in the scope of the work towards this Ph.D. thesis, which is why maximum overlap can be recommended generally for best fusion quality performance. However, it should be noted that the number N_p of patches to be processed increases *quadratically* as N_L^{PO} increases linearly. Therefore, the parameter N_L^{PO} trades fusion fidelity versus compu-

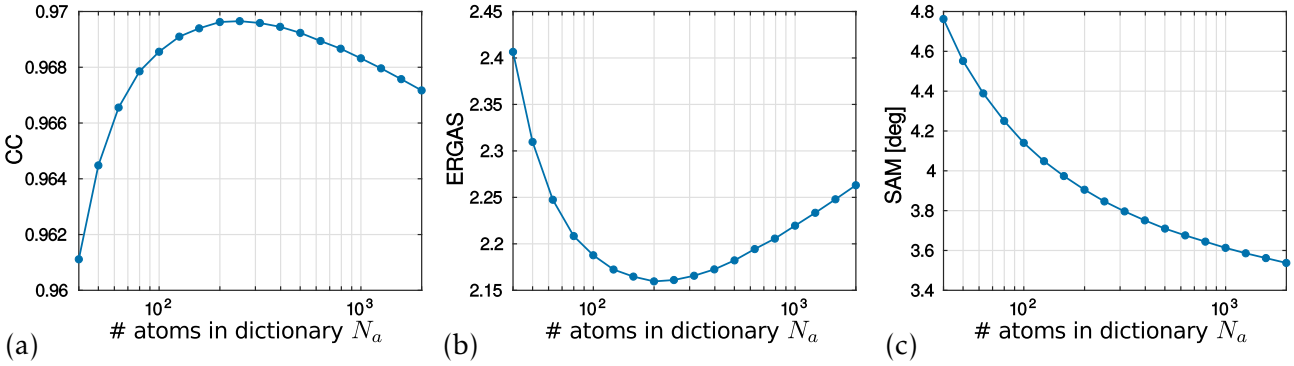


Fig. 46. Analysis of the impact of changes in the dictionary size N_a on the overall fusion quality, which is assessed using (a) CC, (b) ERGAS and (c) SAM as quantitative assessment metrics. Except for N_a, all system settings are as described in the caption of Fig. 42.

tational cost. Finally, even though not explicitly analyzed in this section, the patch size N_{L}^{P} has been empirically found to be best set to 5^2, 6^2 or 7^2. These numbers have been suggested also in related studies [Li and Yang, 2011; Jiang et al., 2012; Zhu and Bamler, 2013; Jiang et al., 2014].

4.4.4 Dictionary Size

The dictionary size N_a is another crucial parameter. Its selection is, again, a trade-off between reconstruction quality and computational cost. Fig. 46 describes the quality metrics CC (left), ERGAS (middle), and SAM (right) as functions of N_a for the JSM-processing of the primary group, i.e. channels 2 (blue) through 5 (red). In this experiment, the patch size is set to $N_{\mathrm{L}}^{P} = 5^2$ with a full patch overlap of $N_{\mathrm{L}}^{PO} = 4$ low-resolution pixels. The regularization parameter is selected as $\lambda' = 1.0$, and the resolution ratio of the input images is $F_{\mathrm{DS}} = 10$. Fig. 46 shows that, as N_a increases, CC and ERGAS improve significantly until they reach peak performance around $N_a = 200$. Beyond that number of atoms, the fusion performance w.r.t. CC and ERGAS gradually declines as N_a further increases. On the other hand, the SAM assessment depicted in Fig. 46 (c) shows continuous spectral fidelity improvement with increasing N_a.

Since the NIR channels capture information less well represented by the atoms contained in the dictionaries than the multispectral channels covered by the Pan sensor SRF, larger N_a numbers have been found to lead to better fusion performance w.r.t. CC and ERGAS. Fig. 44 (b) illustrates the processing time (measured using 128 cores) as a function of N_a. As can be expected, the computational cost increases more than linearly with increasing numbers of dictionary atoms. In the experiments presented later in this chapter, N_p is set to 200 for channels number 1 through 6 and 1,000 for the JSM processing of the two NIR channels.

4.4.5 Summary

Similar to other sparse representation-based pan-sharpening methods, the J-SparseFI algorithm requires a set of parameters to be analyzed and set. Both the quality of the fusion products and the processing time depend on regularization parameters, patch size, patch overlap, and dictionary size. This section provided thorough analyses and recommendations of how to best set these parameters. This study does not aim at providing the best reconstruction results, but a reasonable trade-off between quality and computational cost. Experimental results using parameters selected according to this study and a comparison to

other pan-sharpening methods will be presented in Section 4.5.

- ◇ Regularization parameters λ' and λ: In order to balance the quality and computational processing costs, a parameter value between 1.0 and 10 is recommended. In Section 4.5, both regularization parameters are set to 1.0. Note that (1) λ' is a dimensionless parameter, (2) all quantities involved in Eq. (96) are normalized and (3) all dictionary atoms have a mean value equal to zero. Thus, this recommendation for λ' is generic and does not depend on the scaling or the dynamic range of the data.
- ◇ Patches should be square and contain between $N_{\mathrm{L}}^{P} = 5^2$ and 7^2 low-resolution – and correspondingly more high-resolution – pixels; The number of overlapping pixels in each direction is recommended to be set to the largest possible number, i.e. $N_{\mathrm{L}}^{PO} = N_{\mathrm{L}}^{P} - 1$, for maximum fusion fidelity. In Section 4.5, this concept is followed by setting $N_{\mathrm{L}}^{P} = 5^2$ and $N_{\mathrm{L}}^{PO} = 4$.
- ◇ The favorable dictionary size lies between $N_a = 200$ and $1,000$. For applications primarily requiring high spectral accuracy or high accuracy in NIR channels, larger numbers of dictionary atoms are recommended. In Section 4.5, N_a is set to 200 for the channels in the visible wavelength range and $1,000$ for the NIR channels.

4.5 Performance Evaluation & Comparison to other Methods

In this section, the performance of the proposed J-SparseFI algorithm is investigated and both visually and quantitatively compared to the original SparseFI algorithm as well as to other conventional methods. All experiments are carried out based on the data sets described in Section 4.2.1 using the parameter setting recommended in Section 4.4.

4.5.1 Visual Comparison

For the sake of compactness, only results produced from the data with the higher resolution ratio of $F_{\mathrm{DS}} = 10$ are presented visually. The quantitative assessment presented in the next subsection is done for both $F_{\mathrm{DS}} = 4$ and 10.

From the two input images depicted in Fig. 34, the high-resolution multispectral image has been produced using J-SparseFI and other established methods. Fig. 47 shows true-color RGB compositions of all fusion results and the reference data on a representative magnified spatial subregion. The "ideal" reference image is shown on the upper left. From left to right and top to bottom, Fig. 34 shows the high-resolution multispectral images reconstructed using the Gram Schmidt (GS) method [Laben and Brower, 2000], the PCA method [Shah et al., 2008], the Adaptive IHS method [Rahmani et al., 2010], the synthetic variable ratio merging method (SVR-MM) [Zhang, 1999], the two-step sparse coding method with patch normalization (PN-TSSC) [Jiang et al., 2014], the additive wavelet luminance proportional (AWLP) method [Otazu et al., 2005], SparseFI [Zhu and Bamler, 2013], and the proposed J-SparseFI algorithm.

Noticeably, compared to most of the conventional pan-sharpening methods, sparse reconstruction-based algorithms, i.e. the PN-TSSC method, the SparseFI algorithm, and the J-SparseFI algorithm, produce visually good results even under the situation of the large resolution ratio of $F_{\mathrm{DS}} = 10$. Furthermore, the spectral (color) distortion they introduce are visually less pronounced than those introduced by, e.g., the Gram Schmidt, PCA, and the

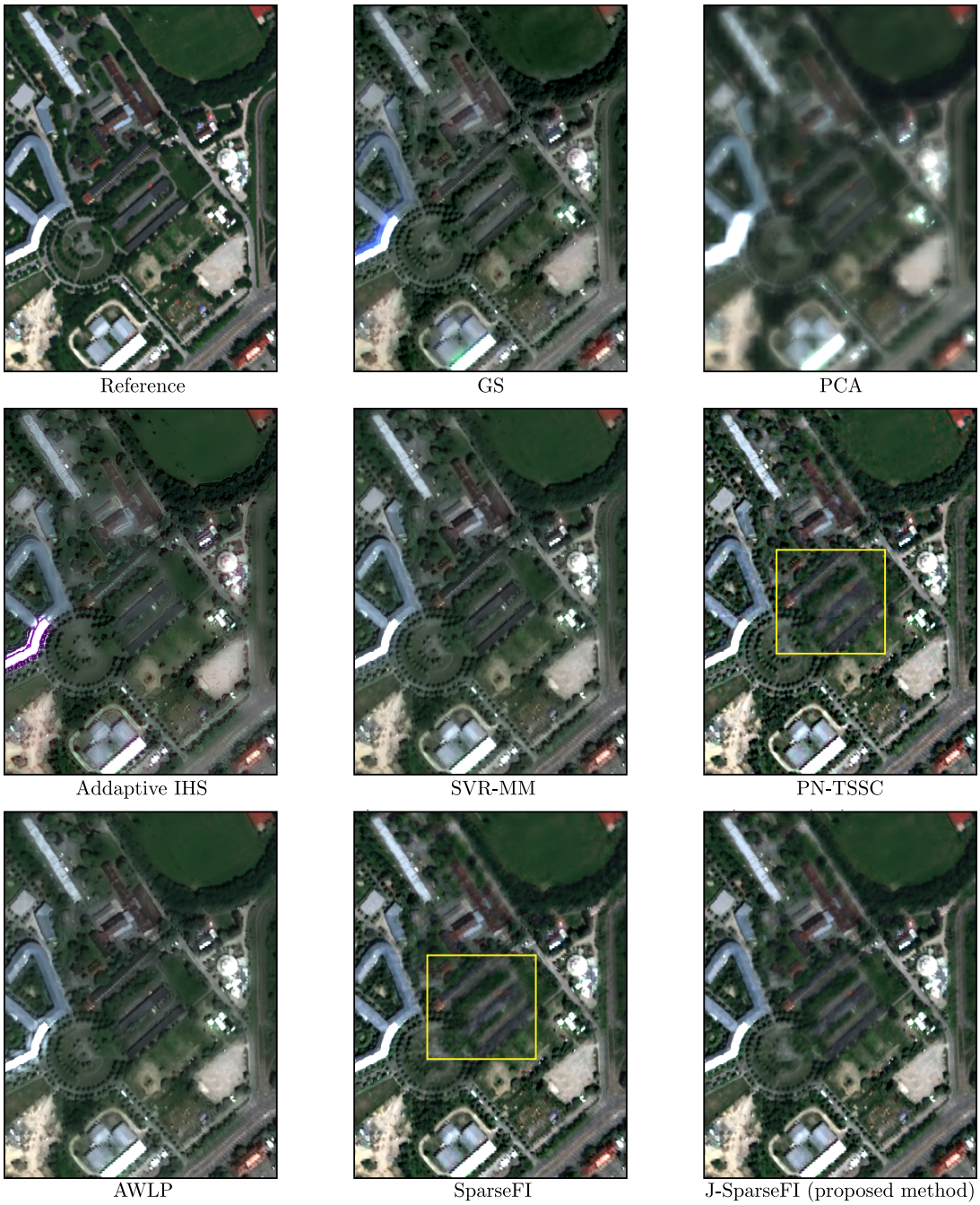


Fig. 47. True-color RGB compositions of high-resolution multispectral reference data (upper left image) and fusion results (remaining images) produced using various pan-sharpening methods including the proposed J-SparseFI algorithm. The underlying input data is the HySpex-based WorldView-2-like imagery with resolution ratio of $F_{\mathrm{DS}} = 10$ described in Section 4.2.1. All images show a representative spatial sub-area of the full processed data set.

adaptive IHS method. Visually, the results produced by SVR-MM and AWLP seem to be similar quality in the selected RGB channels.

Among the three sparse reconstruction-based algorithms, both PN-TSSC and SparseFI tend to introduce artifacts in the reconstructed image, which may be caused by the high uncertainty of the coefficient estimates due to independent reconstructions of the multispectral channels. This effect can be observed in the area marked by the yellow box. Some color distortion and blurred structures are visible in the results produced by PN-TSSC and SparseFI, which are corrected in the results of J-SparseFI. These observations are a first means of con-

		GS	PCA	AIHS	SVR-MM	PN-TSSC	AWLP	SparseFI	J-SparseFI
CC	Coastal	0.9407	0.9321	0.9701	0.9822	0.9834	0.9884	0.98891	**0.98978**
	Blue	0.9447	0.9269	0.9753	0.9858	0.9870	0.9905	0.9907	**0.9911**
	Green	0.9520	0.9035	0.9787	0.9893	0.9900	0.99250	0.99248	**0.9926**
	Yellow	0.9536	0.9197	0.9806	0.9909	0.9910	0.9937	0.99386	**0.99387**
	Red	0.9499	0.9306	0.9799	0.9900	0.9895	0.9929	0.9931	**0.9932**
	Red Edge	0.9509	0.7406	0.9739	0.9905	0.9895	**0.9934**	0.9930	0.9931
	NIR-1	0.9138	0.9363	0.9593	0.9770	0.9763	0.9828	0.9823	**0.9859**
	NIR-2	0.9143	0.9346	0.9591	0.9760	0.9753	0.9821	0.9817	**0.9851**
	Avg.	0.9400	0.9030	0.9721	0.9852	0.9852	0.98955	0.98951	**0.9906**
ERGAS	Coastal	6.9493	7.1860	4.8466	3.8443	3.5856	2.9962	2.9371	**2.8186**
	Blue	7.6585	8.4451	5.1302	4.0857	3.6214	3.0768	3.0566	**2.9908**
	Green	6.6401	8.9791	4.6652	3.4819	2.9684	2.5539	2.5724	**2.5492**
	Yellow	6.8427	8.5181	4.5933	3.3687	2.9369	2.4184	**2.4033**	2.4044
	Red	7.5199	8.4403	4.8849	3.6716	3.3649	2.7354	2.7044	**2.6773**
	Red Edge	3.8131	8.7958	3.3287	1.8152	1.7825	**1.4093**	1.4581	1.4495
	NIR-1	5.6494	10.0878	4.3099	3.0371	3.0386	2.5636	2.6060	**2.3276**
	NIR-2	5.6318	10.0911	4.3072	3.0970	3.0928	2.6185	2.6506	**2.3945**
	Total	6.4477	8.8624	4.5379	3.3640	3.0965	2.5911	2.5889	**2.4897**
SAM		8.0812	7.8150	4.4855	4.0399	4.0404	3.6978	3.2334	**3.0758**

Table 3. Quantitative quality assessment of the fusion results produced from the HySpex-based WorldView-2 input data with a resolution ratio of $F_{\mathrm{DS}} = 4$ introduced in Section 4.2.1. The underlying CC, ERGAS and SAM evaluation metrics are introduced in Section 4.4.1.

firmation that joint reconstruction of highly correlated channels, like J-SparseFI does, is a promising tool to enhancing the robustness of the sparse coefficient estimation step and to reducing the risk of introducing artifacts.

4.5.2 Quantitative Assessment

As for quantitative assessment, the reconstructed high-resolution results are analyzed using the assessment metrics introduced in Section 4.4.1. The first results discussed here are those produced from the input data with a resolution ratio of $F_{\mathrm{DS}} = 4$, which are representative for most of the current optical multi-sensor satellites such as Spot-7, Pléiades, GeoEye-1 and WorldView-2 through 4. Tab. 3 summarizes a quantitative quality assessment of all fusion results. Specifically, evaluation was conducted based on CC (both channel-wise and averaged), ERGAS (both channel-wise and as total metric), and SAM in degrees. The best value in each line is highlighted in bold while the second best value is marked using underlines. Since the average CC values, the total ERGAS values, and SAM indicate overall fusion performance, the corresponding lines are additionally highlighted by a gray background color in the table.

From the table, it can be observed that the pan-sharpening performance strongly depends on the sensor spectral response and/or channel mutual correlation. In general, the table reveals that, perhaps except for GS and PCA, all methods perform well in the primary group of adjacent mutually highly correlated channels, whose spectral ranges match well with the one of the Pan image, namely, the blue, green, yellow, and red channels. Also the coastal channel, which is highly correlated to this group, is reconstructed similarly well. Except for PCA, all methods achieve a better reconstruction in the red-edge channel than in the NIR-1/2 channels. This result can be expected, considering that the SRFs of the NIR spectral channels are not covered by the Pan sensor SRF (see Fig. 33) and that the Pan image is least correlated to those two band images (see Fig. 41).

Comparing the performance of the different methods in this experiment reveals that AWLP,

		GS	PCA	AIHS	SVR-MM	PN-TSSC	AWLP	SparseFI	J-SparseFI
CC	Coastal	0.9002	0.8330	0.9192	0.9258	0.9537	0.9543	0.9553	**0.9588**
	Blue	0.9083	0.8288	0.9264	0.9480	0.9609	0.9617	0.9613	**0.9627**
	Green	0.9325	0.8280	0.9364	0.9659	0.9722	**0.9742**	0.9736	0.9739
	Yellow	0.9316	0.8329	0.9339	0.9691	0.9705	0.97326	0.97317	**0.97328**
	Red	0.9224	0.8374	0.9296	0.9663	0.9642	0.9673	0.9677	**0.9684**
	Red-Edge	0.9500	0.6803	0.9169	**0.9698**	0.9694	0.9697	0.9637	0.9644
	NIR-1	0.8541	0.6838	0.8699	0.9225	0.9224	0.9102	0.8986	**0.9239**
	NIR-2	0.8558	0.6827	0.8712	0.92233	0.92231	0.9109	0.9000	**0.9242**
	Avg.	0.9069	0.7759	0.9129	0.9487	0.9545	0.9527	0.9492	**0.9562**
ERGAS	Coastal	3.5609	4.4695	3.2934	2.9824	2.3969	2.3743	2.3313	**2.2444**
	Blue	3.9223	5.1244	3.6783	2.9663	2.5580	2.4531	2.4661	**2.4262**
	Green	3.2826	4.7761	3.2238	2.3861	2.0747	**1.8866**	1.9212	1.9140
	Yellow	3.3755	4.8882	3.5109	2.3899	2.2058	**1.9854**	1.9938	1.9961
	Red	3.7249	5.1323	3.7532	2.6289	2.5475	2.3194	2.3057	**2.2861**
	Red Edge	2.0131	3.7090	2.3382	1.2200	1.2250	**1.2039**	1.3337	1.3152
	NIR-1	3.0372	4.6334	2.9173	2.1585	2.1649	2.3011	2.4389	**2.1350**
	NIR-2	3.0200	4.6296	2.9011	2.1595	2.1637	2.2900	2.4204	**2.1298**
	Total	3.2884	4.6896	3.2322	2.4192	2.2028	2.1367	2.1819	**2.0803**
SAM		9.4510°	11.7817°	7.7655°	7.3608°	7.3630°	8.1524°	7.1108°	**6.6949°**

Table 4. Quantitative quality assessment of the fusion results produced from the HySpex-based WorldView-2 input data with a resolution ratio of $F_{DS} = 10$ introduced in Section 4.2.1. The underlying CC, ERGAS and SAM evaluation metrics are introduced in Section 4.4.1.

SparseFI and J-SparseFI quantitatively produce the overall best results w.r.t. all assessment metrics, followed by PN-TSSC and SVR-MM. GS and PCA produce the most severe spectral distortions. From this table, it is evident that – among the three sparse representation-based methods – the J-SparseFI algorithm outperforms PN-TSSC and SparseFI, and therefore all other methods under comparison, in almost all assessment metrics. These results confirm the performance improvement achievable through a thorough consideration of (1) the signal correlation between different multispectral channels, (2) the sensor spectral responses and (3) an adaptive selection of the source image for dictionary learning.

The second data set, which involves the uncommonly large resolution ratio of $F_{DS} = 10$, serves as a more challenging scenario to test the limits of the pan-sharpening algorithms under comparison. Tab. 4 summarizes the corresponding quantitative assessment results. As for the overall numbers and highlighted best-ranking algorithms, similar conclusions can be drawn to those described above. Different is that SVR-MM and AWLP are more competitive at this higher discrepancy between the spatial resolution of the multispectral and the Pan input images. J-SparseFI, again, achieves the best overall results in all metrics and also in most of the individual channels. Particular attention should be given to the two NIR bands, which, also in this scenario, are clearly reconstructed best by J-SparseFI. This fact confirms the strength of the proposed concept to not directly use the Pan image for the enhancement of bands which are uncorrelated to the Pan image.

Moreover, the performance superiority of J-SparseFI over PN-TSSC and SparseFI is more pronounced in the case of $F_{DS} = 10$. This outcome is explainable considering the higher discrepancy between the atoms in the coupled low- and high-resolution dictionaries. The estimated sparse coefficient vectors link corresponding high- and low-resolution atoms and are assumed to equally determine representations on the low- and high-resolution level. However, since the coefficients are computed only on the low-resolution level, deficient estimations are more likely to occur at larger resolution ratios. The JSM makes the estimation more robust and less fault-prone compared to individual coefficient reconstructions.

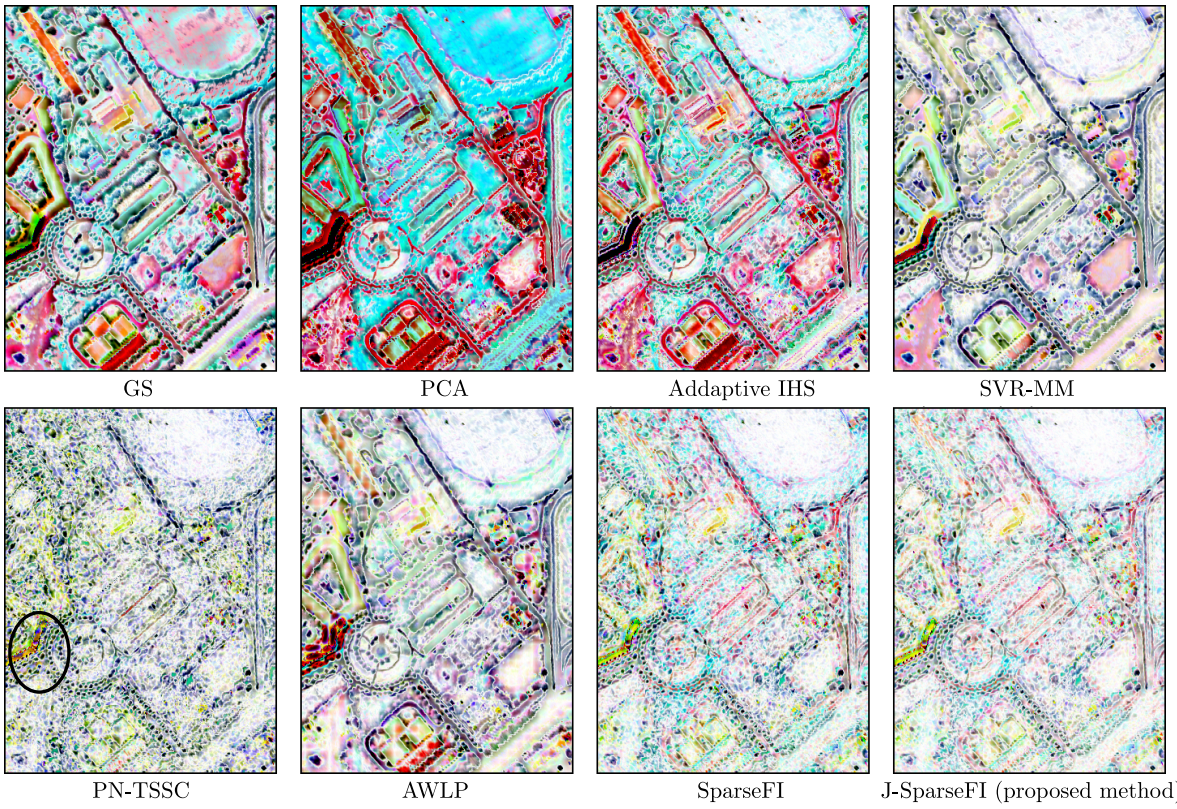


Fig. 48. Difference images between the high-resolution multispectral reference data and pan-sharpening results shown in Fig. 47. All images are false-color compositions with R-G-B being selected as NIR-1, yellow, and coastal, respectively. The relative intensity values are inverted in order to enhance visual interpretability. Hence, white color indicates zero difference while intensive red, blue, and green colors indicate significant errors in the NIR-1, red, and coastal channels, respectively.

4.5.3 Difference Images

The above-discussed means of quantitative evaluation indicates *global* performance assessment. In order to be able to better analyze the spatially varying *local* fusion performance, difference images between each reconstructed high-resolution image and the reference image are created and depicted in Fig. 48. In order to emphasize the quality differences between the different fusion products, the shown results correspond to the data set characterized by $F_{\mathrm{DS}} = 10$. All images are false-color compositions with R-G-B being selected as NIR-1, yellow, and coastal, respectively. Those channels are selected for the following reasons. First, in J-SparseFI, the first (coastal) and seventh (NIR-1) WorldView-2 channels are pan-sharpened using the previously pan-sharpened high-resolution image bands number 2 (blue) and 6 (red edge), respectively, instead of the Pan image. Therefore, it is reasonable to assess the error reduction in those two channels. Second, channel number 4 (yellow) was additionally chosen, because it represents the primary group of channels (i.e., bands number 2 through 5) jointly sharpened via JSM. In Fig. 48, the relative intensity values are inverted for the sake of visualization enhancement. Thus, white color indicates zero difference while intensive red, blue, and green colors indicate significant errors in the NIR-1, red, and coastal channels, respectively. Abrupt color jumps between white and other colors indicate resolution loss or severe color distortions of distinct objects. If no resolution loss is introduced, the difference map should appear as randomly scattered colors. If resolution loss occurs, such abrupt changes will follow spatial object features in the image. In particular, the wider the transition region, the more severe the resolution loss. Based on these rules of thumb, the following conclusions can be drawn.

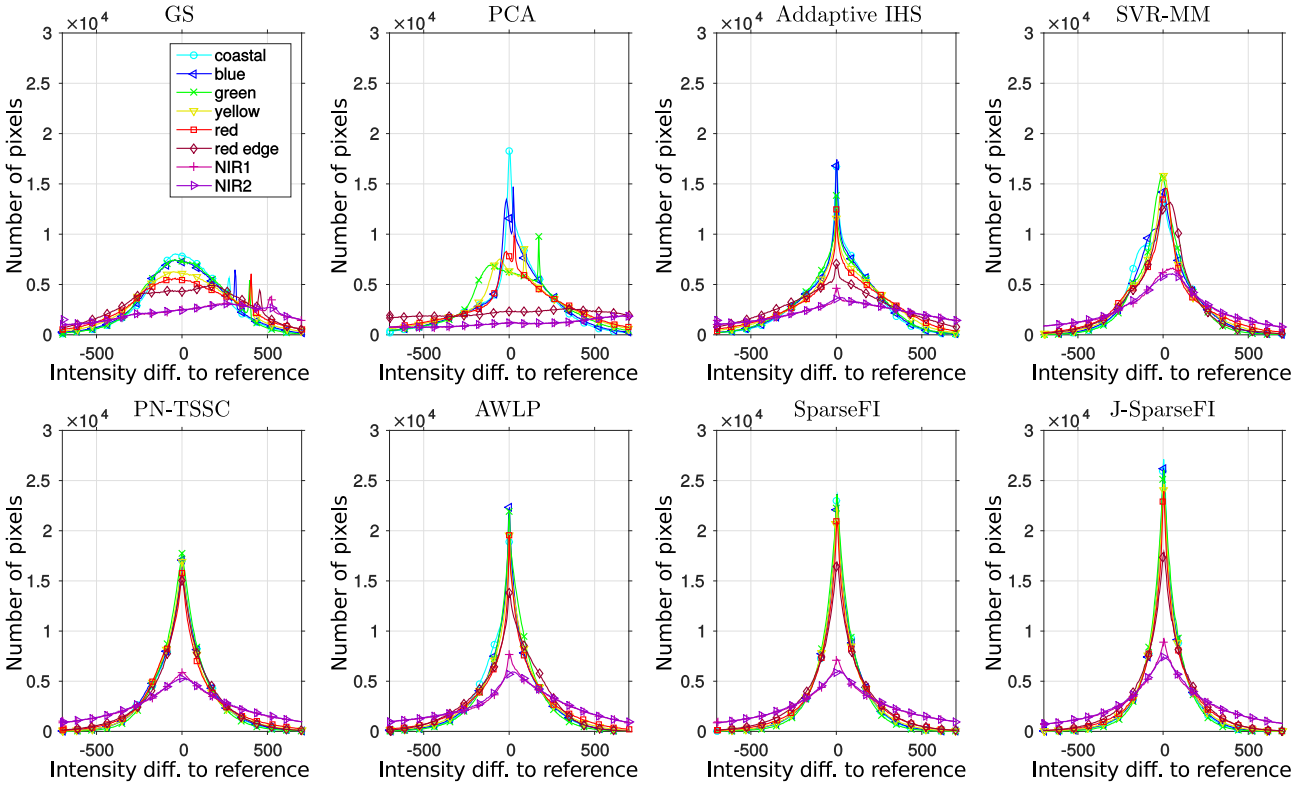
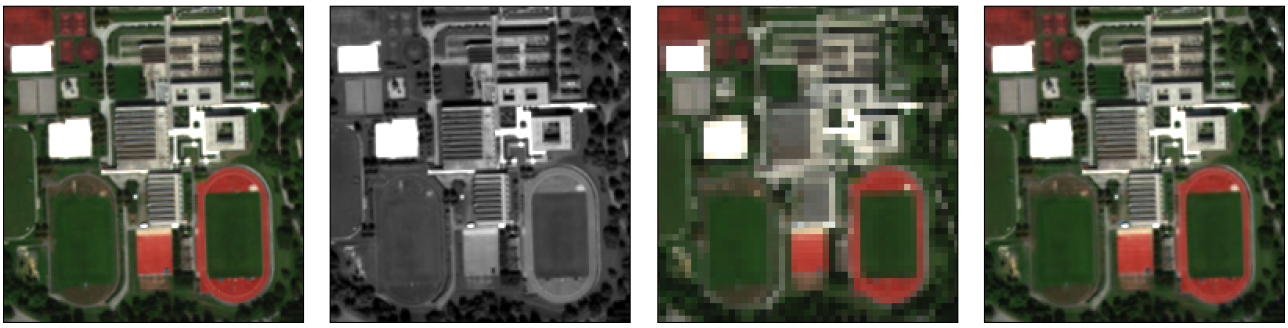


Fig. 49. Histograms of difference images between the pan-sharpening results and original high-resolution multispectral reference data, which are partly depicted in Fig. 47 but calculated on the full spatial scene of $3,600 \times 900$ high-resolution pixels. The underlying input data is the HySpex-based WorldView-2-like imagery synthesized with a resolution ratio of $F_{\mathrm{DS}} = 10$. Each of the eight pan-sharpening products corresponds to one eight-channel difference images. The histograms are plotted separately for every spectral channel. Both input and fusion output data are given as unsigned 16-bit integers.

- ◇ Clear structures can be observed from the difference images. This means that all methods introduce a certain individual, locally varying level of resolution loss relative to the "ideal" reference image. The transition regions visible in the products corresponding to the PN-TSSC, SparseFI and J-SparseFI sparse representation-based fusion methods are narrower than those visible in the images produced by the other methods. This observation indicates better spatial resolution enhancement capacities of the former class of methods.
- ◇ In homogeneous areas – consider, for instance, the sports field located in the upper right corner of the images shown in Fig. 48 – both SparseFI and J-SparseFI produce stable and accurate reconstruction results in all selected channels. The PN-TSSC product, on the other hand, shows a large number of outliers randomly scattered in colors other than white.
- ◇ In terms of spectral distortion, GS and PCA clearly perform the worst, which is indicated by intensive colors dominating the corresponding difference images. As for the other methods, spectral distortion mainly occurs at the boundaries of objects, which can be associated with resolution loss. Particular local performance differences occur in the area marked by a black oval in the lower-left image in Fig. 48. Within that area, the sparse reconstruction-based algorithms most evidently outperform the other methods. All in all, the above-described analysis of the difference images yields the following performance ranking of these three methods in descending order of fusion quality: J-SparseFI, SparseFI, and PN-TSSC.

After analyzing the local fusion performance exemplarily in a selected area, the full difference images in all bands are used to derive statistical values, which will show to give



Multispectral / High resolution (Reference data, 2 m GSD) | Panchromatic / Hich resolution (input data, 2 m GSD) | Multispectral / Low resolution (input data, 8 m GSD) | J-SparseFI product (fusion result, 2 m GSD)

Fig. 50. True-color RGB compositions of a representative area extracted from the original 2-m WorldView-2 reference image shown in Fig. 36 (left), the input high-resolution Pan image (center left), the input low-resolution multispectral image (center right), and the pan-sharpened high-resolution multispectral image produced using the proposed J-SparseFI method (right).

further rise about overall fusion strengths and weaknesses. Fig. 49 illustrates histograms corresponding to differences between the reconstructed high-resolution images and the reference image for the full data sets shown in Fig. 34. As was mentioned in Section 4.2.1, the data are given in the unsigned 16-bit integer domain.

Error distributions are readily interpretable from the histograms. First, the fact that all histograms are centered around zero indicates that none of the methods introduces systematic global brightening and darkening effects. Second, it can be observed that the sparse reconstruction-based methods, as well as AWLP, introduce less reconstruction errors in total. Among those four methods, the histograms corresponding to J-SparseFI are characterized by both nearly perfect symmetry and the highest peaks centered at zero intensity difference. Thus, J-SparseFI evidently produces the most perfectly or nearly perfectly reconstructed pixels. Moreover, all methods individually perform worst in the NIR-1 and NIR-2 channels. This can be expected considering their comparatively low correlation with the other channels and, in particular, the Pan channel. The most significant variations among the histograms of all methods is visible in the red-edge channel. Considering that red-edge corresponds to the most spectrally dynamic transition zone in the VNIR wavelength range, which connects the visible and the NIR spectral regions, performance variation can be expected to be most distinctive in that channel.

4.6 Experiment on Real WorldView-2 Data

As a final practical demonstration, experiments have been carried out using the real WorldView-2 imagery introduced in Section 4.2.2. From left to right, Fig. 50 depicts (1) the original 2-m WorldView-2 reference image in the selected area shown in Fig. 36, (2) the input high-resolution Pan image, (3) the input low-resolution multispectral image, and (4) the high-resolution multispectral image reconstructed using the proposed J-SparseFI pan-sharpening method, respectively.

Tab. 5 shows how this reconstructed full-scene high-resolution multispectral image quantitatively compares to the methods used for the comparative analyses in the previous section. This experiment on real data seconds the conclusions drawn from the previous experiments: Compared to established and previously proposed sparse representation-based pan-sharpening methods, the sophisticated proposed J-SparseFI fusion algorithm produces high-resolution imagery which is of competitive to superior quality and is, thus, well suited

		GS	PCA	AIHS	SVR-MM	PN-TSSC	AWLP	SparseFI	J-SparseFI
CC	Coastal	0.9496	0.9219	0.9514	0.9868	0.9835	0.9901	0.9903	**0.9909**
	Blue	0.9516	0.9316	0.9602	0.9905	0.9876	0.9925	0.9926	**0.9930**
	Green	0.9541	0.9651	0.9673	0.9956	0.9929	0.9967	0.9967	**0.9967**
	Yellow	0.9566	0.9520	0.9660	0.9953	0.9920	0.9964	0.9965	**0.9965**
	Red	0.9557	0.9411	0.9683	0.9936	0.9903	0.9957	0.9958	**0.9959**
	Red Edge	0.9203	0.8735	0.9423	0.9526	0.9598	**0.9721**	0.9709	0.9712
	NIR-1	0.8890	0.8950	0.9201	0.9014	0.9133	0.9428	0.9461	**0.9534**
	NIR-2	0.8899	0.8976	0.9181	0.8996	0.9117	0.9434	0.9469	**0.9539**
	Avg.	0.9334	0.9222	0.9492	0.9644	0.9664	0.9787	0.9795	**0.9814**
ERGAS	Coastal	8.1778	15.3905	7.7619	4.7480	4.7153	3.6393	3.6143	**3.4992**
	Blue	6.6915	12.9799	5.9557	3.4549	3.4214	2.6526	2.6500	**2.5770**
	Green	5.3672	10.9706	4.8622	2.1965	2.1527	1.4618	1.4734	**1.4534**
	Yellow	5.3977	11.2271	5.0998	2.3167	2.3617	1.5741	**1.5599**	1.5612
	Red	5.9154	12.0958	5.1742	2.8609	2.8348	1.8658	1.8468	**1.8325**
	Red Edge	4.4649	6.0002	4.0125	3.1831	2.9479	**2.4371**	2.4992	2.4806
	NIR-1	5.5526	5.1504	4.7257	4.9232	4.6007	3.7359	3.6278	**3.3835**
	NIR-2	5.3852	4.9673	4.6493	4.8789	4.5476	3.6348	3.5245	**3.2909**
	Total	5.9619	10.5194	5.3873	3.7254	3.5834	2.7723	2.7370	**2.6274**
SAM		5.7766	7.1671	4.7357	4.6045	4.6051	4.0006	3.6920	**3.5471**

Table 5. Quantitative quality assessment of the fusion results produced from the real WorldView-2 data introduced in Section 4.2.2. The underlying CC, ERGAS and SAM evaluation metrics are introduced in Section 4.4.1.

for operational processing of multiresolution multispectral-Pan data acquired by current and future Earth observation missions.

5

The J-SparseFI-HM Algorithm – A Solution to the Hyperspectral-Multispectral Data Fusion Problem

5.1 Introduction

This chapter presents the Jointly Sparse Fusion of Hyperspectral and Multispectral Imagery (J-SparseFI-HM) algorithm, which has been designed in the scope of this dissertation to solve the inherently ill-posed problem of fusing hyper- and multispectral remote sensing data of different spectral and spatial resolution to obtain a high-resolution hyperspectral data set.

For the developed methodology, some of the key elements of the previously discussed J-SparseFI pan-sharpening algorithm have been adopted and further developed. Those elements include sparse representation-based reconstruction of image patches, collective processing of mutually correlated channels via joint sparse recovery and the construction of coupled low- and high-resolution dictionary pairs from the input high-resolution image.

In comparison to the pan-sharpening setup, the sensor fusion problem discussed in this chapter implicates potentially much larger spectral dimensions. The corresponding rich spectral information is comprehensively utilized in J-SparseFI-HM by a fully automatic, locally adaptive correlation-based hyperspectral grouping (CorHySpeG) module, which has been designed to identify, for each patch-reconstruction task, partially overlapping groups of hyperspectral channels that are found optimal for collective processing.

Depending on the sensor combination subject to fusion, large parts of the operational wavelength range of the hyperspectral instrument may lie outside of the spectral portions covered by the multispectral imager and are, therefore, not even indirectly represented by the high-resolution input data. This challenge is met, firstly, by finding, for each patch and channel group, the optimum combination of multispectral channels via non-negative linear regression, secondly, by exploiting the spatial relationship (quantified by PSFs) between the input low-resolution and estimated high-resolution hyperspectral data, and, thirdly, by minimizing the local representation error via sparse recovery of non-locally distributed patches.

In contrast to J-SparseFI, J-SparseFI-HM conducts sparse recovery on the low- and high-resolution level simultaneously. That is, instead of using, solely, the low-resolution dictionary and corresponding measurements for the sparse coefficient estimation, both the low- and high-resolution coupled dictionaries are used. High-resolution measurements are extracted from the reconstruction result generated in the previous iteration as will be further explained below. This new feature has a significantly robustifying impact on the representation accuracy of the recovered coefficients. Moreover, new atom selection methods have been developed and tested yielding fusion quality improvements by a replacement of the previously used nearest-neighbor selection approach by a random process.

Another novelty of J-SparseFI-HM is the procedure for acquiring the mean value information induced in the produced patches. In J-SparseFI, the channel-specific mean value of a patch is assumed to be equal to the mean value of the corresponding input low-resolution patch. Since this assumption potentially introduces an approximation error, it is not adopted in J-SparseFI-HM but replaced by an adaptive estimation procedure.

In the spatial domain, J-SparseFI-HM conducts multi-stage processing in an alternating manner. Patch-based reconstruction is implemented in one module combining local with non-local information retrieval, whereas a global processing module optimizes on the full spatial extend of the high-resolution input data to enhance spatial consistency especially across individually processed patches.

A thorough, fully parallelized implementation of J-SparseFI-HM allows processing of large-scale Earth observation data in a short period of time.

A comprehensive experimental setup has been compiled to elaborately analyse the performance of the proposed fusion scheme as well as its competitiveness against the state of the art. In addition to statistical quality assessment, an application-driven investigation on the accuracy gain in material classification through hyperspectral resolution enhancement further validates the effectiveness of the proposed data fusion methodology.

The remainder of this chapter is structured as follows: In Section 5.2, the developed methodology is described, including the underlying sensor observation model and the two aforementioned processing modules. The experimental setup is detailed in Section 5.3. Section 5.4 presents comprehensive system and parameter analyses. Final, the quality of the fusion products obtained via J-SparseFI-HM and the state of the art is assessed and compared for a variety of sensor combinations in Section 5.5.

5.2 Methodology

5.2.1 Overview: The Alternating Local – Non-local – Global Fusion Approach

The fusion framework developed in the scope of this thesis is an alternating multi-stage procedure. Fig. 51 shows a flowchart of the J-SparseFI-HM algorithm.

The aim is to produce a high-resolution hyperspectral image $\boldsymbol{I}_Z$ given a low-resolution hyperspectral image $\boldsymbol{I}_Y$ and a higher-resolution multispectral image $\boldsymbol{I}_X$ as inputs.

In the spatial domain, the input images are fused both patch-wise and on a full spatial scale in an alternating manner. The patch-wise processing part constitutes the first module of J-SparseFI-HM, whereas the full-image processing part is implemented in a second module.

In Fig. 51, those two modules are framed in green and indicated as "Local + Non-local Processing Module" and "Global Processing Module", respectively.

The local–non-local module commences by producing estimates for the target high-spatial high-spectral resolution image patches using local statistics of both input data sets. Subsequently, patches are refined via joint sparse recovery using coupled dictionary pairs which constitute non-local patch information. All patches are typically set to be largely overlapping in order to mitigate potential discontinuities at patch boundaries. After being individually processed, all patches are averaged at the overlapping regions to form an intermediate fusion product $\boldsymbol{I}_{\hat{Z}}$.

In order to further improve the overall spatial consistency, in particular at patch transitions, the second module performs global error minimization while incorporating point spread function (PSF)-based matching of the fusion product $\boldsymbol{I}_{\hat{Z}}$ to the original data.

In the spectral domain, the fusion process is conducted also both on a local and global spectral scale. Here, "local" means that groups of adjacent, highly correlated hyperspectral channels are sharpened jointly, whereas the term "global" refers to the simultaneous processing of all channels. The group-wise processing scheme is integrated in the first module whereas the collective processing of all channels is implicated in the second module. On the one hand, group-wise processing enables generating and using highly wavelength-specific dictionaries; on the other hand, it allows robust patch reconstruction via joint sparse recovery of highly correlated channels. The subsequent collective processing of all channels enhances the product's overall spectral consistency via spectral response function (SRF)-based model matching.

The algorithm's primary input comprises the mere low-resolution hyperspectral and higher-resolution multispectral data. Additionally, the sensors' SRFs and PSFs are required. If not available *a priori*, those quantities can be estimated directly from the two input images. Moreover, the local–non-local process requires as input a low-pass filtered and downsampled version $\boldsymbol{I}_{X,\mathrm{L}}$ of the multispectral input data. Furthermore, the two modules are inter-connected by the hyperspectral data $\boldsymbol{I}_{\hat{Z}}$ which is both an input and the product of each module. In particular, $\boldsymbol{I}_{\hat{Z}}$ is alternatingly updated until convergence. As will be shown in Section 5.4.4, the convergence rate primarily depends on the first initial image $\boldsymbol{I}_{\hat{Z}}$, which may be, for instance, an interpolated and subsampled version of the input low-resolution hyperspectral image or the product of another fusion method.

The remainder of this section provides a step-by-step description of the developed methodology. In particular, Section 5.2.2 presents the observation model underlying J-SparseFI-HM, Section 5.2.3 describes the local–non-local processing module, the global optimization module is described in Section 5.2.4 and the procedure for estimating sensor information from the data is presented in Section 5.2.5.

5.2.2 System Model and Sensor Relationship

This section presents the observation model underlying J-SparseFI-HM. This model describes and links the most crucial system parameters including spectral and spatial characteristics of the input and synthesized sensors. Recall Fig. 27 on page 41, which depicts an illustration of this multi-sensor relationship.

In order to focus on the actual data fusion task, the model employed in this thesis is based on the following assumptions: First, both input images are pre-processed up to the same

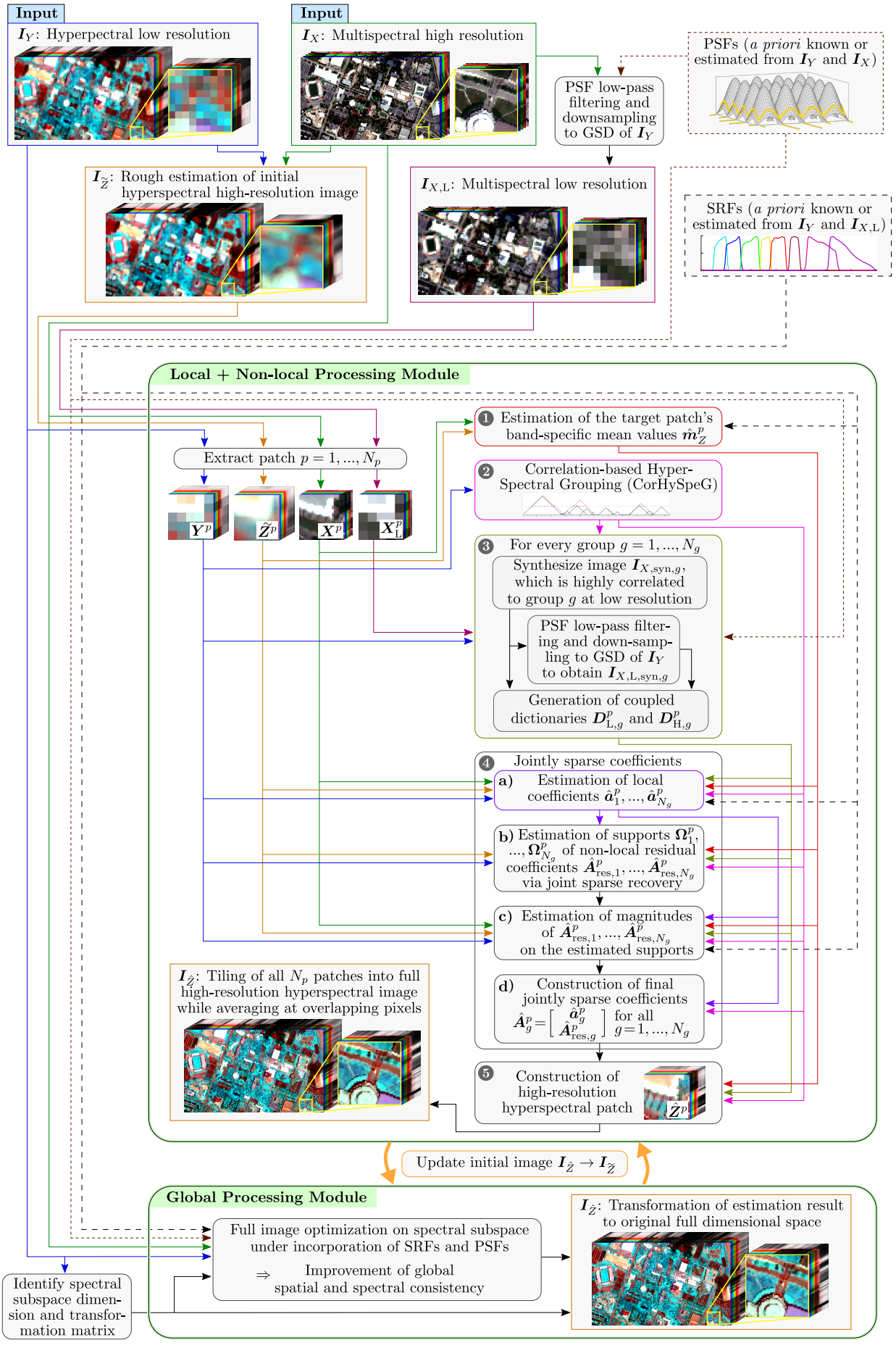


Fig. 51. Flowchart of the J-SparseFI-HM algorithm.

product level, i.e. DNs, radiance or reflectance values; Second, the input images cover the same geographic area and are spatially co-registered.

The inter-sensor relationship incorporated in J-SparseFI-HM has been implicitly introduced in Section 2.2.3 on pages 42 through 45, where the simulation procedure of the input low-spectral resolution data $\boldsymbol{I}_X$ and low-spatial resolution data $\boldsymbol{I}_Y$ from original high-spatial high-spectral resolution imagery $\boldsymbol{I}_O$ is described. Here, the inverse problem of finding $\boldsymbol{I}_Z \approx \boldsymbol{I}_O$ is addressed given $\boldsymbol{I}_X$ and $\boldsymbol{I}_Y$. The fusion product $\boldsymbol{I}_Z$ will be constructed such that its ground sampling distance (GSD) matches and its PSFs well-approximate the respective quantities corresponding to the high-spatial resolution input image $\boldsymbol{I}_X$. Similarly, the spectral responsivity – quantified by SRFs – of the synthesized sensor Z is assumed to be similar to the channel-specific spectral responsivity of the input hyperspectral sensor Y.

Accounting for the above-mentioned aspects and assumptions, the observation model linking the input sensors X and Y to the synthesized sensor Z can be directly derived from Eqs. (48)–(49) on page 45, which simplify to the following system of equations:

$$\boldsymbol{I}_X = \boldsymbol{R}\boldsymbol{I}_Z + \boldsymbol{E}_X \tag{99}$$
$$\boldsymbol{I}_Y = \boldsymbol{I}_Z\boldsymbol{S} + \boldsymbol{E}_Y \tag{100}$$

In Eqs. (99) and (100), the quantities $\boldsymbol{R} := \boldsymbol{R}^{Y,X}$ and $\boldsymbol{S} := \boldsymbol{S}^{X,Y}$ are the – not necessarily normalized – sparse spectral and spatial transformation matrices defined in Eqs. (44) and (45), which respectively hold information about the sensors' SRFs and PSFs. The terms $\boldsymbol{E}_X$ and $\boldsymbol{E}_Y$ represent system errors which are assumed to be approximable by independent zero-mean Gaussian random processes. This model was proposed by Eismann and Hardie [2004] and has been used as sensor observation model in image fusion since then (see e.g. Yokoya et al. [2012]; Simões et al. [2015]).

5.2.3 Local + Non-local Processing Module

This section provides a step-by-step description of the local–non-local processing module. The process starts by dividing each of the input images $\boldsymbol{I}_X$, $\boldsymbol{I}_Y$, $\boldsymbol{I}_{X,\mathrm{L}}$ and $\boldsymbol{I}_{\tilde{Z}}$ into N_p patches $\boldsymbol{X}^p$, $\boldsymbol{Y}^p$, $\boldsymbol{X}_{\mathrm{L}}^p$ and $\tilde{\boldsymbol{Z}}^p$, where, for each $p = 1,\ldots,N_p$, those four patches correspond to the same area on the ground. The patches should be set square and partially overlapping. The patch size can be defined by the total number N_{L}^P ($\sqrt{N_{\mathrm{L}}^P} \in \mathbb{N}$) of low-resolution pixels while the level of overlapping can be set by the number $N_{\mathrm{L}}^{PO} \in \{0,1,\ldots,\sqrt{N_{\mathrm{L}}^P}-1\}$ of low-resolution pixels counted individually along each spatial image dimension. For instance, a patch overlap of $N_{\mathrm{L}}^{PO} = \sqrt{N_{\mathrm{L}}^P} - 1$ implicates that two horizontally or vertically adjacent patches effectively overlap at a total of $N_{\mathrm{L}}^{PO}\sqrt{N_{\mathrm{L}}^P}$ low-resolution pixels. The high-resolution patches $\boldsymbol{X}^p$ and $\tilde{\boldsymbol{Z}}^p$ both contain $N_{\mathrm{H}}^P = F_{DS}^2 N_{\mathrm{L}}^P$ pixels in each band, where F_{DS} denotes the ratio between the ground sampling distances (GSD) of both input sensors, and overlap with any adjacent patch at $N_{\mathrm{H}}^{PO} = F_{DS} N_{\mathrm{L}}^{PO}$ pixels along one spatial direction.

Each of the N_p high-resolution hyperspectral target patches $\hat{\boldsymbol{Z}}^p$ is generated independently from the other patches following steps ❶ through ❺ of the procedure depicted in Fig. 51. Those steps are successively described in the correspondingly labeled subsections further below in this section. But before going into the detailed sequential description, the overall idea is provided in a top-down manner, so that individually estimated and derived quantities can be directly brought into context.

The aim is to obtain $\hat{\boldsymbol{Z}}^p \in \mathbb{R}^{N_b^Y \times N_\mathrm{H}^P}$ as the sum of a band-wise zero-mean estimate $\hat{\boldsymbol{Z}}_0^p \in \mathbb{R}^{N_b^Y \times N_\mathrm{H}^P}$ and band-specific mean values $\boldsymbol{m}_Z^p \in \mathbb{R}^{N_b^Y}$ as follows:

$$\hat{\boldsymbol{Z}}^p = \hat{\boldsymbol{Z}}_0^p + \hat{\boldsymbol{m}}_Z^p \mathbf{1}^{1 \times N_\mathrm{H}^P} \tag{101}$$

In Eq. (101), $\mathbf{1}^{M \times N}$ denotes an $M \times N$-dimensional matrix (in this case, a transposed vector), whose elements are all equal to one. Moreover, $\hat{\boldsymbol{m}}_Z^p \in \mathbb{R}^{N_b^Y}$ contains estimates of the band-specific mean values of the target hyperspectral image patch $\hat{\boldsymbol{Z}}^p \in \mathbb{R}^{N_b^Y \times N_\mathrm{H}^P}$ as will be further described in step ❶ below. The subtlety of the J-SparseFI-HM algorithm lies in the calculation of $\hat{\boldsymbol{Z}}_0^p$, which involves

- ◇ spectral channel grouping (step ❷),
- ◇ generation of group-specific coupled low- and high-resolution dictionary pairs $(\boldsymbol{D}_{\mathrm{L},1}^p, \boldsymbol{D}_{\mathrm{H},1}^p), \ldots, (\boldsymbol{D}_{\mathrm{L},N_g}^p, \boldsymbol{D}_{\mathrm{H},N_g}^p)$ each of which is composed of zero-mean atoms (step ❸), and
- ◇ estimation of group-specific matrices $\hat{\boldsymbol{A}}_{\mathrm{H},1}^p, \ldots, \hat{\boldsymbol{A}}_{\mathrm{H},N_g}^p$ each of which is composed of jointly sparse coefficient vectors (step ❹).

In particular, $\hat{\boldsymbol{Z}}_0^p$ is obtained by computing

$$\hat{\boldsymbol{Z}}_0^p = \widehat{\boldsymbol{P}}^p \left(\widehat{\boldsymbol{D}}_\mathrm{H}^p \widehat{\boldsymbol{A}}^p \right)^T \tag{102}$$

where $\widehat{\boldsymbol{D}}_\mathrm{H}^p$ comprises N_g high-resolution dictionaries (one for each spectral group),

$$\widehat{\boldsymbol{D}}_\mathrm{H}^p = \left[\boldsymbol{D}_{\mathrm{H},1}^p, \ldots, \boldsymbol{D}_{\mathrm{H},N_g}^p \right], \tag{103}$$

$\widehat{\boldsymbol{A}}^p$ is a block-diagonal matrix built from the estimated jointly sparse coefficient vectors,

$$\widehat{\boldsymbol{A}}^p = \mathrm{diag}\left(\hat{\boldsymbol{A}}_1^p, \ldots, \hat{\boldsymbol{A}}_{N_g}^p \right), \tag{104}$$

and $\widehat{\boldsymbol{P}}^p$ is a sparse, fat matrix, which projects the possibly overlapping, not necessarily equally sized N_g groups of hyperspectral channels into a regular hyperspectral representation of N_b^Y channels. Overlapping channels are averaged. The matrix $\widehat{\boldsymbol{P}}^p$ is a side-product of the spectral channel grouping process conducted in step ❷ and is structured like the following schematic matrix:

$$\widehat{\boldsymbol{P}}^p = \begin{bmatrix} \begin{bmatrix} 1 & \ddots & 0 \\ 0 & & 1 \end{bmatrix} \leftarrow \text{group } 1 & & 0 \\ & \begin{bmatrix} 1 & 1 & & 0 \\ & & \ddots & \\ 0 & & & 1 \end{bmatrix} \leftarrow \text{group } 2 & \\ & & \ddots & \\ 0 & & \text{group } N_g \rightarrow & \begin{bmatrix} 1 & \ddots & 0 \\ 0 & & 1 \end{bmatrix} \end{bmatrix} \Big\} N_b^Y \text{ rows} \tag{105}$$

The exact definition of $\widehat{\boldsymbol{P}}^p$ will be give later in Eq. (107) after the spectral grouping concept has been introduced.

With this introduction given, the following subsections successively describe steps ❶ through ❺ of the local–non-local processing module of J-SparseFI-HM.

❶ Estimation of the Target Patch's Band-Specific Mean Values $\boldsymbol{m}_Z^p$

Since the low-resolution hyperspectral input image $\boldsymbol{I}_Y$ is assumed to be an approximation of a low-pass filtered and down-sampled version of the target image $\boldsymbol{I}_{\hat{Z}}$, any low-resolution pixel in $\boldsymbol{I}_Y$ can be assumed to be approximable by a weighted sum of high-resolution pixels in $\boldsymbol{I}_{\hat{Z}}$ located in a neighborhood of that low-resolution pixel. Hence, the mean value $\hat{m}_{Z,j}^p$ of any patch $\hat{\boldsymbol{Z}}^p$ in any channel $j = 1,\ldots,N_b^Y$ is likely to *deviate* from the mean value $m_{Y,j}^p$ of the corresponding low-resolution hyperspectral patch $\boldsymbol{Y}^p$ in the same band, because $m_{Y,j}^p$ incorporates information from high-resolution pixels in $\boldsymbol{I}_{\hat{Z}}$ that are located *outside of the target patch* $\hat{\boldsymbol{Z}}^p$. Therefore, the estimation of $\boldsymbol{m}_Z^p$ should be based on the high-resolution input patches available, namely $\widetilde{\boldsymbol{Z}}^p$ and $\boldsymbol{X}^p$. With a view to the sensor observation model stated in Eqs. (99) and (100), the high-resolution input data $\boldsymbol{I}_X$ and synthesized data $\boldsymbol{I}_Z$ are linked via the matrix $\boldsymbol{R}$ which is constructed from the input sensors' SRFs. Thus, $\boldsymbol{R}$ is used here to access the mean-value information of $\boldsymbol{X}^p$ for the estimation of $\boldsymbol{m}_Z^p$ as follows:

$$\hat{\boldsymbol{m}}_Z^p = \min_{\boldsymbol{m}_Z^p} \left\{ \frac{1}{N_b^Y} \left\| \boldsymbol{m}_Z^p - \frac{1}{N_{\mathrm{H}}^P} \hat{\boldsymbol{Z}}^p \mathbf{1}^{N_{\mathrm{H}}^P \times 1} \right\|_2^2 + \frac{\mu_X}{N_b^X} \left\| \boldsymbol{R}\boldsymbol{m}_Z^p - \frac{1}{N_{\mathrm{H}}^P} \boldsymbol{X}^p \mathbf{1}^{N_{\mathrm{H}}^P \times 1} \right\|_2^2 \right\} \tag{106}$$

In Eq. (106), the parameter μ_X weighs the relative influence of $\boldsymbol{X}^p$ on the estimation result. Similarly, this parameter will be used in steps ❹ a) through c) to trade the influence of $\boldsymbol{X}^p$ against the influence of the other input patches $\boldsymbol{Y}^p$ and $\widetilde{\boldsymbol{Z}}^p$ on the estimation of $\widehat{\boldsymbol{A}}^p$.

❷ Correlation-Based HyperSpectral Grouping (CorHySpeG)

This section describes the Correlation-based HyperSpectral Grouping (CorHySpeG) concept developed for and implemented in the local–non-local processing module of J-SparseFI-HM.

The idea is to increase the patch reconstruction accuracy and robustness by collectively processing mutually correlated channels via joint sparse recovery. Moreover, processing individual channel subsets allows for adapting system quantities, such as dictionaries, to the corresponding wavelength ranges.

Given a hyperspectral image patch $\boldsymbol{Y}^p \in \mathbb{R}^{N_{\mathrm{L}}^P \times N_b^Y}$ composed of N_b^Y spectral channels, the aim is to find a channel grouping which meets the following requirements:

(1) Any pair of channels in the same group has a minimum mutual correlation of $\theta \in \mathbb{R}$, where θ is a sensor-specific, relatively high correlation value.
(2) Every channel is contained in at least one group.
(3) Groups consist of consecutive channels (usually corresponding to center wavelengths in ascending order).
(4) Two adjacent groups share ideally close to – but no more than – 50 % of the channels in the smaller of both groups.

The rationale behind requirement (4) is that, on the one hand, allowing adjacent groups to overlap while averaging those high-resolution channels that have been produced as parts of multiple groups increases spectral reconstruction fidelity. One the other hand, restricting the number of overlapping channels to half the size of the smaller of two adjacent groups obviates an unnecessary overhead of computational costs.

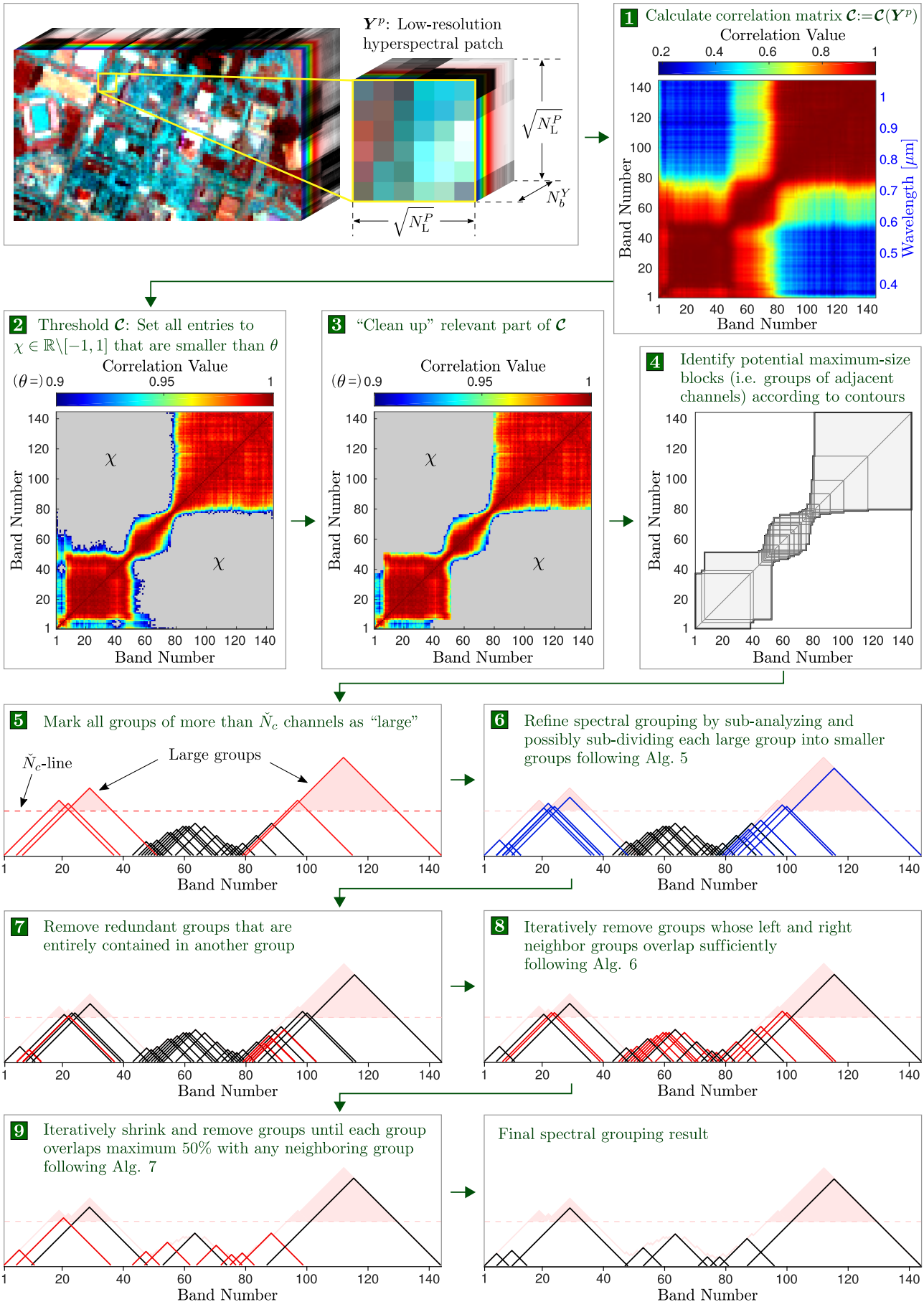


Fig. 52. Flowchart illustrating the Correlation-based HyperSpectral Grouping (CorHySpeG) concept detailed in Alg. 4.

According to requirement (3), a spectral grouping result can be fully described by the following three quantities:

- $N_g \in \mathbb{N}$: Number of channel groups.
- $\boldsymbol{n}_c \in \mathbb{N}^{N_g}$: Numbers of channels per group.
- $\boldsymbol{h} \in \mathbb{N}^{N_g}$: Indices of the leading channels.

For instance, element $h_i \in \{1,\dots,N_b^Y\}$ indexes the leading channel in group number $i \in \{1,\dots,N_g\}$. More specifically, the i-th group is composed of $n_{c,i}$ hyperspectral channels whose indices are the following: $\{h_i, h_i+1,\dots,h_i+n_{c,i}-1\}$.

The CorHySpeG algorithm is described in detail in Alg. 4 and illustrated in Fig. 52. All nine steps of Alg. 4 are individually illustrated in the correspondingly labeled fields in Fig. 52.

In steps **1** through **4**, an initial spectral grouping is calculated that satisfies requirements (1) through (3).

In order to increase the algorithm's robustness and reduce its sensitivity to variation in θ, groups, which are initially found as "large" (by means of their numbers $n_{c,i}$ of channels being larger than a pre-defined group size $\check{N}_c$) are identified in step **5** and possibly split into sub-groups in step **6**. The latter procedure is detailed in Alg. 5 and illustrated in Fig. 53. The idea is to refine the initial spectral grouping result by locally adjusting θ. It is reasonable not to jointly sharpen excessively many channels, as that would not be likely to lead to higher reconstruction accuracy compared to the joint processing of moderately many channels. The number $\check{N}_c$ of channels, above which groups are considered as "large", is a less critical system parameter which can be set according to the spectral sampling distance. For common airborne and space-based spectrometers, numbers between 20 and 40 for $\check{N}_c$ have

Algorithm 4 Correlation-based Hyperspectral Grouping (CorHySpeG)

Input:
- $\boldsymbol{Y} := \boldsymbol{Y}^p \in \mathbb{R}^{N_L^P \times N_b^Y}$, $p \in \{1,\dots,N_p\}$: Local hyperspectral patch
- $\theta \in [-1,1] \subset \mathbb{R}$: Minimum correlation between channels in any spectral group
- $\check{N}_c \in \mathbb{N}$: Number of channels above which groups are examined and possibly subdivided

1: Calculate initial overall correlation matrix $\boldsymbol{\mathcal{C}} := \boldsymbol{\mathcal{C}}(\boldsymbol{Y}^p) \in \mathbb{R}^{N_b^Y \times N_b^Y}$ of current patch $\boldsymbol{Y}^p$;
2: Threshold $\boldsymbol{\mathcal{C}}$: For all $i,j \in \{1,\dots,N_b^Y\}$ with $\boldsymbol{\mathcal{C}}_{i,j} < \theta$ set $\boldsymbol{\mathcal{C}}_{i,j} = \chi$, where $\chi \in \mathbb{R}\backslash[-1,1]$ is arbitrary;
3: "Clean up" $\boldsymbol{\mathcal{C}}$: Set all entries $\boldsymbol{\mathcal{C}}_{i,j}$ and $\boldsymbol{\mathcal{C}}_{j,i}$ to χ for which there is a $(k,l) \in \{(k',l') \in \mathbb{N}\times\mathbb{N} \,|\, j < k' < i, l' = j\} \cup \{(k',l') \in \mathbb{N}\times\mathbb{N} \,|\, k' = i, i < l' < j\}$ s.t. $\boldsymbol{\mathcal{C}}_{k,l} = \chi$;
4: Identify all potential maximum-size θ-*blocks*† and calculate initial values of the three output quantities N_g, $\boldsymbol{h}$ and $\boldsymbol{n}_c$;
5: Identify and mark all groups that are composed of more than $\check{N}_c$ bands as "large";
6: For each large group conduct spectral sub-grouping following Alg. 5 and embed resulting quantities $N_{g,\text{sub}}$, $\boldsymbol{h}_{\text{sub}}$ and $\boldsymbol{n}_{c,\text{sub}}$ into overall grouping by updating N_g, $\boldsymbol{h}$ and $\boldsymbol{n}_c$;
7: Remove redundant groups that are entirely contained in another group; Update N_g, $\boldsymbol{h}$ and $\boldsymbol{n}_c$;
8: Iteratively remove groups whose left and right neighbor groups overlap sufficiently following Alg. 6; Update N_g, $\boldsymbol{h}$ and $\boldsymbol{n}_c$;
9: Iteratively shrink groups until each group overlaps maximum 50% with any of its neighbor groups following Alg. 7; Update N_g, $\boldsymbol{h}$ and $\boldsymbol{n}_c$;

Output:
- $N_g \in \mathbb{N}$: Number of channel groups.
- $\boldsymbol{n}_c \in \mathbb{N}^{N_g}$: Number of channels per group.
- $\boldsymbol{h} \in \mathbb{N}^{N_g}$: Indices of the leading channels.

† Define θ-*block* – or simply *block* – as a square sub-matrix, centered along the diagonal, whose elements are all greater than or equal to θ; A block in a correlation matrix corresponds to a group of adjacent channels. Note, that θ-*blocks* have been introduced first in Alg. 3 on page 86.

been found to lead to fusion performance improvements as will be discussed in the experimental part of this chapter.

Once all large groups have been analysed and potentially sub-divided, the individual sub-grouping results are merged into the overall grouping. This step is realized by updating N_g, $\boldsymbol{n}_c$ and $\boldsymbol{h}$.

As a result of step 6, groups may have been generated that are entirely contained in other groups. Those redundant groups are identified and removed in step 7.

Finally, steps 8 and 9 iteratively eliminate and shrink groups until requirement (4) is met. Alg. 6 describes the iterative group elimination process conducted in step 8. The idea is to remove any group $i \in \{1,\dots,N_g\}$ from the grouping, whose "left" and "right" neighboring groups – indexed by $i-1$ and $i+1$, respectively – sufficiently overlap without the center group i. More precisely, group i is eliminated if groups $i-1$ and $i+1$ share at least 50 % of the channels of the smaller of both groups. In every iteration, the algorithm finds and eliminats the group i whose neighbours share the highest number of channels relative to the number of channels in the smaller of both groups. This process terminates when there is no pair of left and right neighboring groups left which share more than 50 % of the smaller group's channels.

In step 9, the remaining groups are iteratively shrinked until no pair of adjacent groups share more than 50 % of the smaller group's channels. The detailed procedure of this step is described in Alg. 7.

With the final spectral grouping result $(N_g, \boldsymbol{n}_c, \boldsymbol{h})$ at hand, side products can be directly derived which will be used in steps 3 and 5 of J-SparseFI-HM. Those side products include the above-mentioned matrix $\widehat{\boldsymbol{P}}^p \in \mathbb{R}^{N_b^Y \times \sum_{i=1}^{N_g} n_{c,i}}$ and matrices $\boldsymbol{P}_i^p \in \mathbb{R}^{n_{c,i} \times N_b^Y}$, $i = 1,\dots,N_g$, which are respectively defined as follows:

$$\widehat{P}^p_{k,l} = \begin{cases} 1\,, & \text{if } (k,l) = \left(h_i + b - 1, \sum_{j=1}^{i-1} n_{c,j} + b - 1\right) \text{ for any } i \in \{1,\dots,N_g\},\ b \in \{1,\dots,n_{c,i}\} \\ 0\,, & \text{otherwise} \end{cases} \tag{107}$$

and

$$P^p_{i,k,l} = \begin{cases} 1\,, & \text{if } l = h_i + k - 1 \\ 0\,, & \text{otherwise} \end{cases} \tag{108}$$

for any $i = 1,\dots,N_g$. The latter matrices $\boldsymbol{P}_i^p$ simply extract a subset of channels, namely those contained in group number i, from a regular N_b^Y-dimensional hyperspectral pixel.

Considering the sparseness and knowledge about the positions of the non-trivial entries, neither $\widehat{\boldsymbol{P}}^p$ nor $\boldsymbol{P}_i^p$ need to be explicitely computed or stored as full matrices by the implemented software. Instead, extraction and averaging of bands should be implemented more efficiently. In this manuscript, their introduction merely facilitates the description of the mathematical model underlying J-SparseFI-HM.

3 Generation of Coupled Dictionaries $(\boldsymbol{D}^p_{\mathrm{L},1}, \boldsymbol{D}^p_{\mathrm{H},1}), \dots, (\boldsymbol{D}^p_{\mathrm{L},N_g}, \boldsymbol{D}^p_{\mathrm{H},N_g})$

As is indicated in field 3 in Fig. 51, there is one coupled dictionary pair $(\boldsymbol{D}^p_{\mathrm{L},i}, \boldsymbol{D}^p_{\mathrm{H},i})$ generated for each group $i = 1,\dots,N_g$ of mutually correlated hyperspectral channels. The dictionary pairs $(\boldsymbol{D}^p_{\mathrm{L},i}, \boldsymbol{D}^p_{\mathrm{H},i}) \in \mathbb{R}^{N^p_{\mathrm{L}} \times N_a} \times \mathbb{R}^{N^p_{\mathrm{H}} \times N_a}$ are called *coupled*, because their atoms,

Algorithm 5 Sub-division of an individual large group (Step 6 in Fig. 52)

Input:
- θ and $\check{N}_c$: Same as in Alg. 4
- $\boldsymbol{\mathcal{C}}_{\text{sub}} \in [\theta,1]^{n_{c,k}\times n_{c,k}}$: Correlation (sub-)matrix corresponding to large group $k \in \{1,\dots,N_g\}$

1: Calculate refined correlation thresholds $\theta_A, \theta_B, \theta_C \in [\theta,1]$ as follows:
- θ_A: Cast one half of the symmetric matrix $\boldsymbol{\mathcal{C}}_{\text{sub}}$ into a vector $\boldsymbol{c}$ and sort it in ascending order. Set $\theta_A = c_m$, where $m = \frac{1}{2}\#\left\{(i,j) \in \{1,\dots,n_{c,k}\}^2 \;\middle|\; |i-j| > \check{N}_c\right\}$;
- θ_B: Set θ_B to a fix value in the upper half of the interval $[\theta,1]$, e.g. $\theta_B = (\theta+2)/3$;
- θ_C: Set $\theta_C = c_{\tilde{m}}$, where $\tilde{m}$ is found s.t. $\sum_{i=1}^{\tilde{m}} c_i \approx \kappa \sum_{i=1}^{M} c_i$ with $M = n_{c,k}(n_{c,k}-1)/2$ being the number of elements in $\boldsymbol{c}$ and $\kappa \in [0,1]$ being a fix small relative integral part, e.g. 1/10;

2: Set $\theta = \min\{\theta_A, \theta_B, \theta_C\}$ and conduct steps 2 through 4 of Alg. 4 to obtain a spectral sub-grouping, i.e. $N_{g,\text{sub}}$, $\boldsymbol{h}_{\text{sub}}$ and $\boldsymbol{n}_{c,\text{sub}}$;

Output:
- $N_{g,\text{sub}} \in \mathbb{N}$: Number of channel sub-groups.
- $\boldsymbol{n}_{c,\text{sub}} \in \mathbb{N}^{N_{g,\text{sub}}}$: Numbers of channels per sub-group.
- $\boldsymbol{h}_{\text{sub}} \in \mathbb{N}^{N_{g,\text{sub}}}$: Indices of the leading channels.

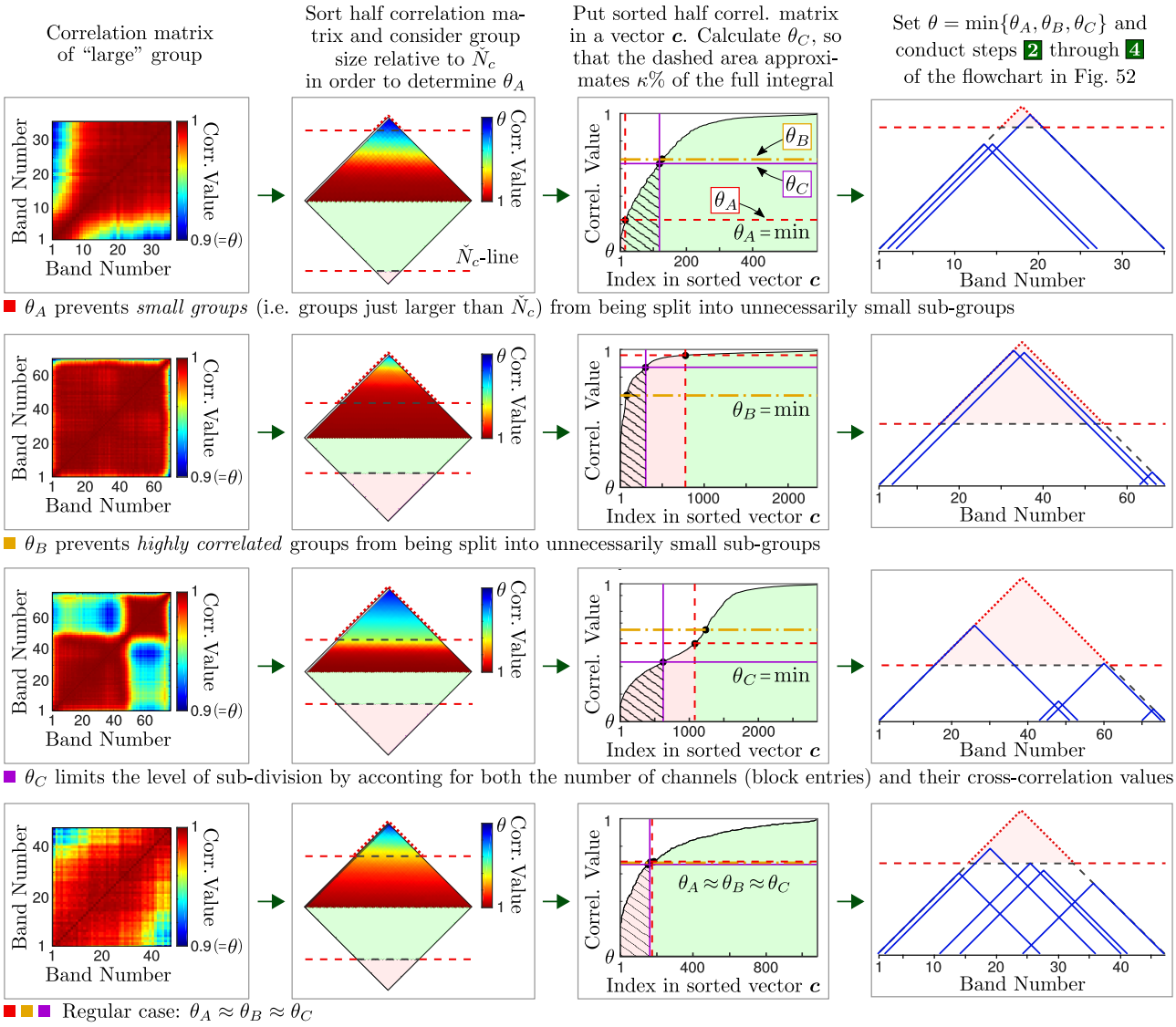


Fig. 53. Illustration of the θ-adjustment and sub-division of large groups described in Alg. 5 (Step 6 in Fig. 52).

Algorithm 6 Iterative removal of redundant center groups (Step 8 in Fig. 52)

Input: N_g, $\boldsymbol{h}$ and $\boldsymbol{n}_c$: As defined in Alg. 4

Calculate the numbers $\boldsymbol{o}_f, \boldsymbol{o}_b \in \mathbb{N}^{N_g}$ of forward and backward overlapping bands between *next-to-adjacent* groups:

$$\boldsymbol{o}_{f,i} = \begin{cases} n_{c,i} - h_{i+2} & \text{, if } i < N_g - 1 \\ 0 & \text{, otherwise} \end{cases}, \qquad \boldsymbol{o}_{b,i} = \begin{cases} 0 & \text{, if } i \leq 2 \\ n_{c,i} - h_{i-2} & \text{, otherwise} \end{cases};$$

while $\exists i \in \{1, \ldots, N_g\}$ s.t. $\left(o_{f,i} > \frac{n_{c,i}}{2} \ \vee \ o_{b,i} > \frac{n_{c,i}}{2}\right)$ **do**

  Set $\hat{i} = \max(\boldsymbol{i}_f)$, where $\boldsymbol{i}_f = \arg\max_{i=1,\ldots,N_g} \left(\left\{\frac{o_{f,i}}{n_{c,i}} \,\middle|\, o_{f,i} > \left\lceil \frac{n_{c,i}}{2} \right\rceil\right\}\right)$ may contain multiple elements;

  Set $\hat{j} = \min(\boldsymbol{i}_b)$, where $\boldsymbol{i}_b = \arg\max_{i=1,\ldots,N_g} \left(\left\{\frac{o_{b,i}}{n_{c,i}} \,\middle|\, o_{b,i} > \left\lceil \frac{n_{c,i}}{2} \right\rceil\right\}\right)$ may contain multiple elements;

  if $\frac{o_{f,\hat{i}}}{n_{c,\hat{i}}} > \frac{o_{b,\hat{j}}}{n_{c,\hat{j}}}$ **then**

    Remove center group $\hat{i}+1$ by updating N_g, $\boldsymbol{h}$ and $\boldsymbol{n}_c$ accordingly;

  else

    Remove center group $\hat{j}-1$ by updating N_g, $\boldsymbol{h}$ and $\boldsymbol{n}_c$ accordingly;

  end if

end while

Output: Updated input quantities N_g, $\boldsymbol{h}$ and $\boldsymbol{n}_c$

Algorithm 7 Iterative shrinkage of strongly overlapping groups (Step 9 in Fig. 52)

Input: N_g, $\boldsymbol{h}$ and $\boldsymbol{n}_c$: As defined in Alg. 4

Calculate the numbers $\boldsymbol{o}_f, \boldsymbol{o}_b \in \mathbb{N}^{N_g}$ of forward and backward overlapping bands between *adjacent* groups:

$$\boldsymbol{o}_{f,i} = \begin{cases} n_{c,i} - h_{i+1} & \text{, if } i < N_g \\ 0 & \text{, otherwise} \end{cases}, \qquad \boldsymbol{o}_{b,i} = \begin{cases} 0 & \text{, if } i = 1 \\ n_{c,i} - h_{i-1} & \text{, otherwise} \end{cases};$$

while $\exists i \in \{1, \ldots, N_g\}$ s.t. $\left(o_{f,i} > \frac{n_{c,i}}{2} \ \vee \ o_{b,i} > \frac{n_{c,i}}{2}\right)$ **do**

  Set $\hat{i} = \max(\boldsymbol{i}_f)$, where $\boldsymbol{i}_f = \arg\max_{i=1,\ldots,N_g} \left(\left\{\frac{o_{f,i}}{n_{c,i}} \,\middle|\, o_{f,i} > \left\lceil \frac{n_{c,i}}{2} \right\rceil\right\}\right)$ may contain multiple elements;

  Set $\hat{j} = \min(\boldsymbol{i}_b)$, where $\boldsymbol{i}_b = \arg\max_{i=1,\ldots,N_g} \left(\left\{\frac{o_{b,i}}{n_{c,i}} \,\middle|\, o_{b,i} > \left\lceil \frac{n_{c,i}}{2} \right\rceil\right\}\right)$ may contain multiple elements;

  if $\frac{o_{f,\hat{i}}}{n_{c,\hat{i}}} > \frac{o_{b,\hat{j}}}{n_{c,\hat{j}}}$ **then**

    if $\hat{i} > 1 \ \wedge \ h_{\hat{i}-1} + n_{c,\hat{i}-1} - h_{\hat{i}+1} = o_{f,\hat{i}} - 1$ **then**

      Remove group $\hat{i}$ by updating N_g, $\boldsymbol{h}$ and $\boldsymbol{n}_c$ accordingly;

    else

      Shrink group $\hat{i}$ by removing its *last* channel, i.e. updating $n_{c,\hat{i}} = n_{c,\hat{i}} - 1$;

    end if

  else

    if $\hat{j} < N_g \ \wedge \ h_{\hat{j}-1} + n_{c,\hat{j}-1} - h_{\hat{j}+1} = o_{f,\hat{j}} - 1$ **then**

      Remove group $\hat{j}$ by updating N_g, $\boldsymbol{h}$ and $\boldsymbol{n}_c$ accordingly;

    else

      Shrink group $\hat{j}$ by removing its *first* channel, i.e. updating $n_{c,\hat{j}} = n_{c,\hat{j}} - 1$ and $h_{\hat{j}} = h_{\hat{j}} + 1$;

    end if

  end if

end while

Output: Updated input quantities N_g, $\boldsymbol{h}$ and $\boldsymbol{n}_c$

$\boldsymbol{d}^p_{\mathrm{L},i,k}$ and $\boldsymbol{d}^p_{\mathrm{H},i,k}$, $k = 1,\ldots,N_a$, correspond to the same 2-D structure. Recall that the above-introduced 2-D image patches $\boldsymbol{X}^p \in \mathbb{R}^{N_b^X \times N_{\mathrm{H}}^P}$ and $\boldsymbol{Y}^p \in \mathbb{R}^{N_b^Y \times N_{\mathrm{L}}^P}$ have their spatial information stored in a single dimension. Similarly, the coupled 1-D dictionary atoms represent 2-D spatial information but at different spatial resolutions and GSDs, which, respectively, correspond to the spatial resolutions and GSDs of the two input image images $\boldsymbol{I}_Y$ and $\boldsymbol{I}_X$. Therefore, at least one data source those coupled dictionary pairs are generated from has to have a spatial resolution at least as high as the resolution of $\boldsymbol{I}_X$. Since the initial image $\boldsymbol{I}_{\tilde{Z}}$ might not be of high resolution, but merely be sampled to the GSD of $\boldsymbol{I}_X$, it is not used for this purpose. There have been attempts in the literature to conduct dictionary training based on *external* high-resolution data, i.e. data other than the algorithm's input, $\boldsymbol{I}_X$ and $\boldsymbol{I}_Y$, [Li and Yang, 2011]. The obvious drawbacks of that approach are, however, the required availability, storage and processing of additional auxiliary data and the fact that external images, generally, do not conform to the nature of the scene under reconstruction.

Inspired by the efficient and effective dictionary generation procedure implemented in J-SparseFI, the coupled dictionaries are generated *locally* and use the *input high-resolution image* $\boldsymbol{I}_X$ as data source. However, the new dictionary learning approach differs from the method designed for J-SparseFI in the following two points:

(1) *Specific High-resolution Data Source*: While the pan-sharpening setup provides only *one* high-resolution channel, namely the panchromatic one, for dictionary generation, the hyperspectral-multispectral fusion setup offers multiple – usually uncorrelated – high-resolution multispectral images. In J-SparseFI, the potential problem of some multispectral channels not being well covered (in terms of SRFs) by or correlated to the panchromatic channel are met by an implementation of the option to utilize previously reconstructed high-resolution multispectral channels as training source instead of the panchromatic one. The latter option is used if one of the reconstructed channels is higher correlated with the bands under reconstruction than the original panchromatic channel is. In J-SparseFI-HM, the analogous problem of hyperspectral channels not being represented well by any of the multispectral channels is even more severe and likely to occur. It is addressed by *synthesizing* – for each group $i = 1,\ldots,N_g$ of mutually correlated hyperspectral channels – a high-resolution channel $\boldsymbol{I}^p_{X,\mathrm{syn},i}$ such that its low-pass filtered and down-sampled version $\boldsymbol{I}^p_{X,\mathrm{L,syn},i}$ locally approximates the *mean* hyperspectral band in group i as best as possible. More precisely, $\boldsymbol{I}^p_{X,\mathrm{syn},i}$ is computed as the linear combination of all high-resolution input multispectral channels, i.e.,

$$\boldsymbol{I}^p_{X,\mathrm{syn},i} = \sum_{b=1}^{N_b^X} \hat{c}_b \boldsymbol{I}_{X,b}\,, \tag{109}$$

where $\boldsymbol{I}^p_{X,b}$ is the b-th band in $\boldsymbol{I}_X$ and the coefficients $\hat{\boldsymbol{c}} = [\hat{c}_1,\ldots,\hat{c}_{N_b^X}]$ are calculated such that the corresponding combination of low-resolution multispectral bands well-approximate the mean hyperspectral band in group i locally in terms of non-negative least squares. That is,

$$\hat{\boldsymbol{c}} = \min_{\boldsymbol{c}} \left\| (\boldsymbol{c}\boldsymbol{X}^p_{\mathrm{L}} - \bar{\boldsymbol{y}}^p_i)^T \right\|_2^2 \qquad \text{s.t. } c_b \geq 0 \text{ for all } b = 1,\ldots,N_b^X \tag{110}$$

where

$$\bar{\boldsymbol{y}}^p_i = n_{c,i}^{-1} \boldsymbol{1}^{1\times n_{c,i}} \boldsymbol{P}^p_i \boldsymbol{Y}^p \tag{111}$$

is the average hyperspectral patch in group i.

Finally, a low-resolution version of $\boldsymbol{I}^p_{X,\mathrm{syn},i}$ is needed for the generation of the low-resolution dictionary $\boldsymbol{D}^p_{\mathrm{L},i}$ alongside $\boldsymbol{D}^p_{\mathrm{H},i}$. For this purpose, it is computationally much more efficient to use the available low-resolution multispectral image $\boldsymbol{I}_{X,\mathrm{L}}$ together with the above-calculated coefficient vector $\boldsymbol{c}$ instead of low-pass filtering and downsampling $\boldsymbol{I}^p_{X,\mathrm{syn},i}$. Thus, it is obtained as follows:

$$\boldsymbol{I}^p_{X,\mathrm{L},\mathrm{syn},i} = \sum_{b=1}^{N_b^X} \hat{c}_b \boldsymbol{I}_{X,\mathrm{L},b}\,. \tag{112}$$

In the scope of the work towards this thesis, the non-negative least squares formulation stated in Eq. (110) was found to consistently lead to higher image reconstruction accuracy than other approaches for calculating $\boldsymbol{c}$ that have been implemented and tested. Those other approaches included (1) non-constrained linear regression and (2) calculation of coefficients which minimize the correlation between $\bar{\boldsymbol{y}}^p_i$ and $\boldsymbol{c}\boldsymbol{X}^p_{\mathrm{L}}$.

(2) *Local Dictionary Atom Selection*: This paragraph addresses the criteria based upon which candidate low- and high-resolution patch pairs $(\boldsymbol{x}^q_{\mathrm{L}}, \boldsymbol{x}^q_{\mathrm{H}}) \in \mathbb{R}^{1\times N^P_{\mathrm{L}}} \times \mathbb{R}^{1\times N^P_{\mathrm{H}}}$, $q \in \{1,\ldots,N_p\}$, that are respectively extracted from $\boldsymbol{I}^p_{X,\mathrm{L},\mathrm{syn},i}$ and $\boldsymbol{I}^p_{X,\mathrm{syn},i}$, are selected for inclusion in the coupled low- and high-resolution dictionary pair $(\boldsymbol{D}^p_{\mathrm{L},i}, \boldsymbol{D}^p_{\mathrm{H},i})$.

In J-SparseFI, the local coupled dictionaries are built up from those patches which are located in a neighborhood of the current patch p under reconstruction. This method is computationally cheap and a natural first approach when changing from a global (SparseFI) to local (J-SparseFI) dictionaries. Further considerations suggest, however, that there may be other approaches to systematically selecting patches, i.e. atoms, that can potentially improve the fusion performance in terms of reconstruction quality.

One way of defining atom selection methods is by associating each method k with a mass function μ_k which quantitatively measures the "suitability" of each of the N_p candidate patches for inclusion in the coupled dictionary pair. In this work, suitability of a candidate patch $q = 1,\ldots,N_p$ is quantified via a comparison of patch q to the current patch $p = 1,\ldots,N_p$ undergoing reconstruction. Subsequently to calculating all values $\mu^{p,q}$, $q = 1,\ldots,N_p$, all candidate patches are sorted, i.e. ranked, according to their μ-values. Finally, the N_a most suitable patches are selected for inclusion in the dictionary, where $N_a < N_p$ denotes the pre-defined dictionary size. N_a should be set such that the underlying linear systems are underdetermined. With a view to Eq. (128) further below, this is the case if $N_a > N^P_{\mathrm{L}} + N^P_{\mathrm{H}}$.

In the following, three atom selection approaches are introduced in addition to the nearest neighbor selection method used in J-SparseFI. All four approaches will be experimentally tested and compared in Section 5.4.

◇ *Nearest Neighbor Patches:* μ_{nn}
Each of the N_p candidate dictionary pairs $(\boldsymbol{x}^q_{\mathrm{L}}, \boldsymbol{x}^q_{\mathrm{H}})$, $q = 1,\ldots N_p$, is associated with 2-D coordinates $\boldsymbol{u} \in \mathbb{N}\times\mathbb{N}$ representing its center location in the 2-D image coordinate system. In order to build up the local dictionary pair $(\boldsymbol{D}^p_{\mathrm{L},i}, \boldsymbol{D}^p_{\mathrm{H},i})$ corresponding to the current patch p under reconstruction, the distance between patch q and patch p is calculated in terms of the s-norm, $1 \le s \le \infty$:

$$\mu^{p,q}_{\mathrm{nn}} = \|\boldsymbol{u}_p - \boldsymbol{u}_q\|_s \tag{113}$$

The N_a closest patches, i.e. those with the *lowest* μ-values, are included in the respective high- and low-resolution dictionaries. In the current J-SparseFI-HM implementation, the maximum norm $\|\cdot\|_{s=\infty}$ is used.

◇ *Most Correlated Patches:* μ_{corr}
This approach aims at finding those N_a patches that are most similar, i.e. correlated, to the current patch p under reconstruction on the high-resolution level. The suitability measure is defined as

$$\mu_{\text{corr}}^{p,q} = \left| \sigma_{\boldsymbol{x}_{\text{H}}^p, \boldsymbol{x}_{\text{H}}^q} \Big/ \left(\sigma_{\boldsymbol{x}_{\text{H}}^p}, \sigma_{\boldsymbol{x}_{\text{H}}^q} \right) \right|, \tag{114}$$

where $\sigma_{\boldsymbol{x}_{\text{H}}^p}$ and $\sigma_{\boldsymbol{x}_{\text{H}}^p, \boldsymbol{x}_{\text{H}}^q}$ denote the standard deviation of $\boldsymbol{x}_{\text{H}}^p$ and the covariance between $\boldsymbol{x}_{\text{H}}^p$ and $\boldsymbol{x}_{\text{H}}^q$, respectively, as defined in Eqs. (28) and (30) on page 32. Note, that the right hand side of Eq. (114) is equal to the absolute correlation between $\boldsymbol{x}_{\text{H}}^p$ and $\boldsymbol{x}_{\text{H}}^q$. Taking the absolute value in Eq. (114) makes patch pairs with high *negative* correlation value equally suitable for inclusion in the dictionary as patches with high positive correlation value. As will become clear from the description of the jointly sparse coefficient estimation (step 4) below, the leading sign is irrelevant in this case, because, on the one hand, selected patches are modified to have zero-mean before inclusion in the dictionary and, on the other hand, the sparse coefficients stored in $\widehat{\boldsymbol{A}}^p$ and multiplied with $\widehat{\boldsymbol{D}}_{\text{H}}^p$ in Eq. (102) can take both positive and negative values. The N_a patches with the *highest* μ-values are selected.

◇ *Least Correlated Patches:* μ_{uncorr}
This approach is based on selecting the *least similar* patches, i.e. those having the lowest absolute correlation to the current patch under reconstruction. More specifically, the dictionary is composed of one atom corresponding to the current patch under reconstruction and $N_a - 1$ atoms that are lowly correlated to the first atom. This concept can be quantified using the following measure:

$$\mu_{\text{uncorr}}^{p,q} = \begin{cases} 0 & \text{if } q = p \\ \left| \sigma_{\boldsymbol{x}_{\text{H}}^p, \boldsymbol{x}_{\text{H}}^q} \Big/ \left(\sigma_{\boldsymbol{x}_{\text{H}}^p}, \sigma_{\boldsymbol{x}_{\text{H}}^q} \right) \right| & \text{if } q \neq p \end{cases} \tag{115}$$

The N_a patches with the *lowest* μ-values are selected.

◇ *Randomly Distributed Patches:* μ_{rand}
This option selects $N_a - 1$ randomly distributed patches along with the one spatially corresponding to the current patch under reconstruction. A corresponding measure could be defined, for instance, by

$$\mu_{\text{rand}}^{p,q} = \begin{cases} 0 & \text{if } q = p \\ f_{\text{rand}}(q,p) & \text{if } q \neq p \end{cases}, \tag{116}$$

where the function $f_{\text{rand}} : \mathbb{N} \times \mathbb{N} \to]0,1[\subset \mathbb{R}$ is a (pseudo) random number generator.

In practice, this approach should obviously be implemented so that only N_a patch-index numbers are generated and there should be no sorting involved. Moreover, it should be ensured that the same patch is not marked twice for selection.

Before patches are selected and stored as atoms in the coupled high- and low-resolution dictionaries, $\boldsymbol{D}_{\text{H},i}^p$ and $\boldsymbol{D}_{\text{L},i}^p$, their mean values are subtracted and the patches are nor-

malized by dividing each element by the l_2-norm of the high-resolution patch. Hence, all dictionary atoms have a mean value equal to zero and the high-resolution dictionary atoms all have an l_2-norm equal to 1. More specifically, if $\{q_1^p,\ldots,q_{N_a}^p\} \subset \{1,\ldots,N_p\}$ denotes the patch index set found via one of the four above-described selection methods, the local coupled dictionary pair corresponding to the p-th patch and i-th group is generated as follows:

$$\boldsymbol{D}_{\mathrm{L},i}^p = \left[\boldsymbol{d}_{\mathrm{L}}^{q_1^p},\ldots,\boldsymbol{d}_{\mathrm{L}}^{q_{N_a}^p}\right] \tag{117}$$

$$\boldsymbol{D}_{\mathrm{H},i}^p = \left[\boldsymbol{d}_{\mathrm{H}}^{q_1^p},\ldots,\boldsymbol{d}_{\mathrm{H}}^{q_{N_a}^p}\right] \tag{118}$$

where, for each $q = q_1^p,\ldots,q_{N_a}^p$,

$$\boldsymbol{d}_{\mathrm{H}}^q = \left(\boldsymbol{x}_{\mathrm{H}}^q - \overline{\boldsymbol{x}}_{\mathrm{H}}^q\right)^T \cdot \left\|\left(\boldsymbol{x}_{\mathrm{H}}^q - \overline{\boldsymbol{x}}_{\mathrm{H}}^q\right)^T\right\|_2^{-1} \tag{119}$$

$$\boldsymbol{d}_{\mathrm{L}}^q = \left(\boldsymbol{x}_{\mathrm{L}}^q - \overline{\boldsymbol{x}}_{\mathrm{L}}^q\right)^T \cdot \left\|\left(\boldsymbol{x}_{\mathrm{H}}^q - \overline{\boldsymbol{x}}_{\mathrm{H}}^q\right)^T\right\|_2^{-1} \tag{120}$$

and

$$\overline{\boldsymbol{x}}_{\mathrm{H}}^q = (N_{\mathrm{H}}^P)^{-1}\boldsymbol{x}_{\mathrm{H}}^q\mathbf{1}^{N_{\mathrm{H}}^P\times 1} \tag{121}$$

$$\overline{\boldsymbol{x}}_{\mathrm{L}}^q = (N_{\mathrm{L}}^P)^{-1}\boldsymbol{x}_{\mathrm{L}}^q\mathbf{1}^{N_{\mathrm{L}}^P\times 1} \tag{122}$$

The reason for using the same divisor in Eqs. (119) and (120) is the essential correspondence between atoms in the coupled dictionary pair when multiplied by the *same coefficient* in the joint sparse recovery step described further below.

Note that all four selection methods are designed in a way that the current patch p undergoing reconstruction is included in the coupled dictionary pair. This inclusion is important as the atom corresponding to the current patch is likely to contribute the most to the reconstruction. Without loss of generality, the *first* atom in the dictionary pair is assumed to correspond to the current patch, i.e. $q_1^p = p$. This technical information will be used later, when the current (i.e. local) patch is accessed as the leading atom in the dictionary. This atom will be used for the estimation of the *local* coefficient whereas the $N_a - 1$ remaining (non-local) atoms form the basis of the *non-local* coefficient estimation process. The latter step can be considered as minimization of the representation error between the current (local) source patch and the local measurements.

Fig. 54 illustrates the spatial distribution of patches selected for inclusion in sample dictionaries of size $N_a = 200$. The underlying dictionary atom selection methods are (a) nearest neighbor patches (b) most correlated patches and (c) least correlated patches. For each method, five patches corresponding to the first five atoms are shown enlarged, where the color of the surrounding patch frames indicate their respective μ-values. Noticeable, the distribution of the lowly correlated patches composing the right-most dictionary in Fig. 54 appears random, especially if compared to the other two displayed dictionaries. As will be experimentally evidenced later, lowly correlated and fully random dictionaries lead, indeed, to similar overall image reconstruction quality. The spatial distribution of the highly correlated dictionary shown in Fig. 54 (b) clearly shows a pattern which follows the street that is partially contained in the current patch.

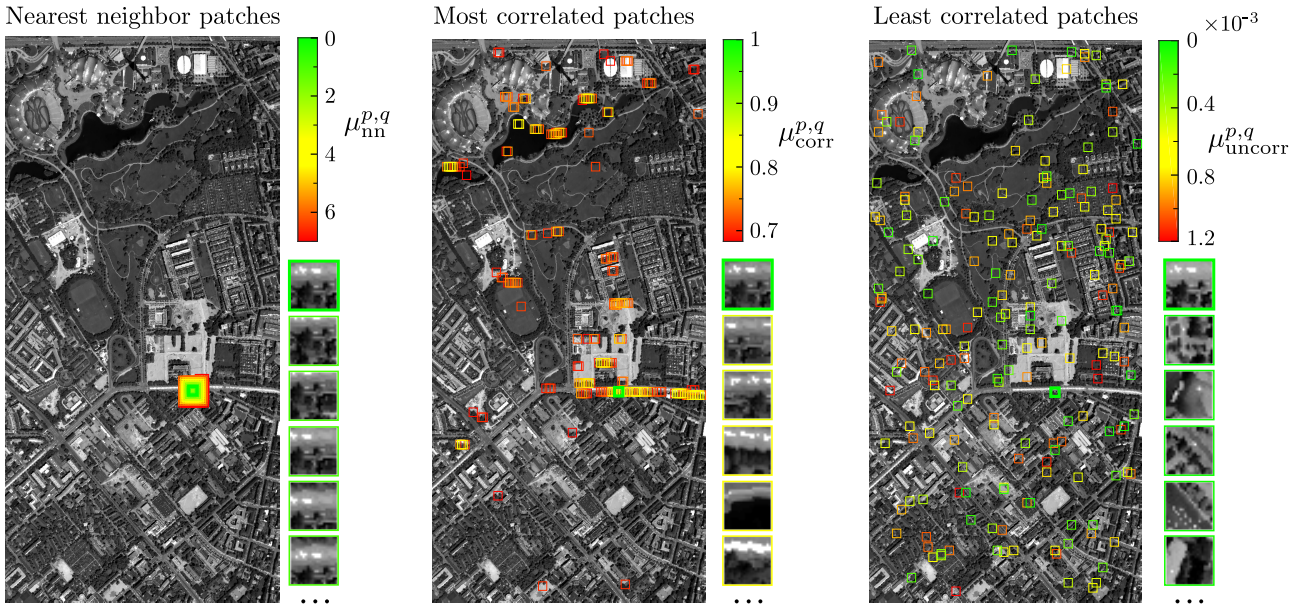


Fig. 54. Spatial distribution of patches selected for inclusion in sample dictionaries of size $N_a = 200$. The underlying dictionary atom selection methods are (a) nearest neighbor patches (b) most correlated patches and (c) least correlated patches. For each method, the patches corresponding to the first five atoms are shown enlarged, where the color of the surrounding patch boundaries indicate their respective μ-values.

④ Estimation of Jointly Sparse Coefficients $\hat{\boldsymbol{A}}_1^p, \ldots, \hat{\boldsymbol{A}}_{N_g}^p$

This section describes the procedure developed for calculating the group-wise jointly sparse coefficients stored in the matrices $\hat{\boldsymbol{A}}_1^p, \ldots, \hat{\boldsymbol{A}}_{N_g}^p$. As indicated in the field marked by ④ in Fig. 51, this procedure comprises the following four successive steps: a) Estimation of *local* coefficients $\hat{\boldsymbol{a}}_i^p \in \mathbb{R}^{1 \times n_{c,i}}$, $i = 1, \ldots, N_g$; Estimation of b) supports and c) magnitudes of *non-local* residual coefficients $\hat{\boldsymbol{A}}_{\mathrm{res},i}^p \in \mathbb{R}^{N_a - 1 \times n_{c,i}}$, $i = 1, \ldots, N_g$; d) Composition of final jointly sparse coefficient vectors $\hat{\boldsymbol{A}}_i^p \in \mathbb{R}^{N_a \times n_{c,i}}$ from $\hat{\boldsymbol{a}}_i^p$ and $\hat{\boldsymbol{A}}_{\mathrm{res},i}^p$.

In order to estimate the contributions of the local- and non-local dictionary atoms separately, the low- and high-resolution dictionaries are temporarily split as follows:

$$\boldsymbol{D}_{\mathrm{L},i}^p = \left[\boldsymbol{d}_{\mathrm{L},i}^p, \boldsymbol{D}_{\mathrm{res,L},i}^p\right] \tag{123}$$

$$\boldsymbol{D}_{\mathrm{H},i}^p = \left[\boldsymbol{d}_{\mathrm{H},i}^p, \boldsymbol{D}_{\mathrm{res,H},i}^p\right] \tag{124}$$

In Eqs. (123) and (124), the atoms $\boldsymbol{d}_{\mathrm{L},i}^p \in \mathbb{R}^{N_{\mathrm{L}}^p}$ and $\boldsymbol{d}_{\mathrm{H},i}^p \in \mathbb{R}^{N_{\mathrm{H}}^p}$ both correspond to the current (local) patch under reconstruction whereas all remaining atoms stored in $\boldsymbol{D}_{\mathrm{res,L},i}^p \in \mathbb{R}^{N_{\mathrm{L}}^p \times N_a - 1}$ and $\boldsymbol{D}_{\mathrm{res,H},i}^p \in \mathbb{R}^{N_{\mathrm{H}}^p \times N_a - 1}$ correspond to non-local patches.

Steps (a) through (c) fully incorporate the sensor observation model stated in Eqs. (99) and (100) above. The coefficients are computed in a way that the product $\widehat{\boldsymbol{D}}_{\mathrm{H}}^p \widehat{\boldsymbol{A}}^p$ and, therefore, the estimated high-resolution hyperspectral patch $\hat{\boldsymbol{Z}}^p$ satisfies the observation model relative to the input patches $\boldsymbol{Y}^p$ and $\boldsymbol{X}^p$ as best as possible while not deviating significantly from the initial patch $\widetilde{\boldsymbol{Z}}^p$. The following paragraphs describe steps a) through d) (see Fig. 51).

a) *Estimation of Local Coefficients* $\hat{\boldsymbol{a}}_1^p, \ldots, \hat{\boldsymbol{a}}_{N_g}^p$

The coefficients $\hat{\boldsymbol{a}}_i^p \in \mathbb{R}^{1\times n_{c,i}}$ determine the contribution of the local atom $\boldsymbol{d}_{\mathrm{H},i}^p$ to the zero-mean patch $\hat{\boldsymbol{Z}}_0^p$ in those bands belonging to group $i \in \{1,\ldots,N_g\}$. The estimation process is conducted via linear regression between the products $\boldsymbol{d}_{\mathrm{L},i}^p \hat{\boldsymbol{a}}_i^p$ and $\boldsymbol{d}_{\mathrm{H},i}^p \hat{\boldsymbol{a}}_i^p$ and the input quantities $\widetilde{\boldsymbol{Z}}^p$, $\boldsymbol{Y}^p$ and $\boldsymbol{X}^p$ whose relative influences to the estimation result are weighted by the system parameters μ_X and μ_Y.

$$\begin{aligned}\left\{\hat{\boldsymbol{a}}_1^p,\ldots,\hat{\boldsymbol{a}}_{N_g}^p\right\} = \min_{\boldsymbol{a}_1^p,\ldots,\boldsymbol{a}_{N_g}^p} \Bigg\{ & \frac{1}{N_{\mathrm{H}}^P N_g}\sum_{i=1}^{N_g}\frac{1}{n_{c,i}}\left\|\left(\boldsymbol{d}_{\mathrm{H},i}^p\boldsymbol{a}_i^p\right)^T + \boldsymbol{P}_i^p\left(\hat{\boldsymbol{m}}_Z^p\mathbf{1}^{1\times N_{\mathrm{H}}^P} - \widetilde{\boldsymbol{Z}}^p\right)\right\|_F^2 \\ & + \frac{\mu_Y}{N_{\mathrm{L}}^P N_g}\sum_{i=1}^{N_g}\frac{1}{n_{c,i}}\left\|\left(\boldsymbol{d}_{\mathrm{L},i}^p\boldsymbol{a}_i^p\right)^T + \boldsymbol{P}_i^p\left(\boldsymbol{m}_Y^p\mathbf{1}^{1\times N_{\mathrm{L}}^P} - \boldsymbol{Y}^p\right)\right\|_F^2 \\ & + \frac{\mu_X}{N_{\mathrm{H}}^P N_b^X}\left\|\boldsymbol{R}\left(\widehat{\boldsymbol{P}}^p\left(\widehat{\boldsymbol{d}}_{\mathrm{H}}^p\widehat{\boldsymbol{a}}^p\right)^T + \hat{\boldsymbol{m}}_Z^p\mathbf{1}^{1\times N_{\mathrm{H}}^P}\right) - \boldsymbol{X}^p\right\|_F^2 \Bigg\}\end{aligned} \tag{125}$$

Similarly to how $\widehat{\boldsymbol{D}}_{\mathrm{H}}^p$ and $\widehat{\boldsymbol{A}}^p$ are defined in Eq. (103) and (104), the single-atom quantities $\widehat{\boldsymbol{d}}_{\mathrm{H}}^p$ and $\widehat{\boldsymbol{a}}^p$ are defined as

$$\widehat{\boldsymbol{d}}_{\mathrm{H}}^p = \left[\boldsymbol{d}_{\mathrm{H},1}^p,\ldots,\boldsymbol{d}_{\mathrm{H},N_g}^p\right] \in \mathbb{R}^{N_{\mathrm{H}}^P\times N_g} \tag{126}$$

and

$$\widehat{\boldsymbol{a}}^p = \mathrm{diag}\left(\boldsymbol{a}_1^p,\ldots,\boldsymbol{a}_{N_g}^p\right) \in \mathbb{R}^{N_g\times\sum_{i=1}^{N_g} n_{c,i}}, \tag{127}$$

respectively.

b) *Estimation of Supports* $\boldsymbol{\Omega}_1^p,\ldots,\boldsymbol{\Omega}_{N_g}^p$ *of Non-local Residual Coefficients* $\hat{\boldsymbol{A}}_{\mathrm{res},1}^p,\ldots,\hat{\boldsymbol{A}}_{\mathrm{res},N_g}^p$

In this step, the supports of the jointly sparse coefficients corresponding to the non-local part of the coupled dictionaries are estimated. Therefore, a first estimate $\hat{\boldsymbol{A}}_{\mathrm{res},i}^{\mathrm{aux},p} \in \mathbb{R}^{N_a-1\times n_{c,i}}$, $i = 1,\ldots,N_g$, is computed using the joint sparse recovery framework introduced in Section 2.3. For each spectral group $i = 1,\ldots,N_g$, the auxiliary jointly sparse coefficient vectors are calculated as follows:

$$\begin{aligned}\hat{\boldsymbol{A}}_{\mathrm{res},i}^{\mathrm{aux},p} = \min_{\boldsymbol{A}_{\mathrm{res},i}^{\mathrm{aux},p}} \Bigg\{ & \frac{\lambda}{\sqrt{n_{c,i}}}\left\|\boldsymbol{A}_{\mathrm{res},i}^{\mathrm{aux},p}\right\|_{1,2} \\ & + \frac{1}{2N_{\mathrm{H}}^P n_{c,i}}\left\|\left(\boldsymbol{d}_{\mathrm{H},i}^p\boldsymbol{a}_i^p + \boldsymbol{D}_{\mathrm{res,H},i}^p\boldsymbol{A}_{\mathrm{res},i}^{\mathrm{aux},p}\right)^T + \boldsymbol{P}_i^p\left(\hat{\boldsymbol{m}}_Z^p\mathbf{1}^{1\times N_{\mathrm{H}}^P} - \widetilde{\boldsymbol{Z}}^p\right)\right\|_F^2 \\ & + \frac{\mu_Y}{2N_{\mathrm{L}}^P n_{c,i}}\left\|\left(\boldsymbol{d}_{\mathrm{L},i}^p\boldsymbol{a}_i^p + \boldsymbol{D}_{\mathrm{res,L},i}^p\boldsymbol{A}_{\mathrm{res},i}^{\mathrm{aux},p}\right)^T + \boldsymbol{P}_i^p\left(\boldsymbol{m}_Y^p\mathbf{1}^{1\times N_{\mathrm{L}}^P} - \boldsymbol{Y}^p\right)\right\|_F^2 \Bigg\}\end{aligned} \tag{128}$$

Just like in Chapter 4, the parameter λ regularizes the sparsity level in $\boldsymbol{A}_{\mathrm{res},i}^{\mathrm{aux},p}$ relative to the model approximation error. The supports,

$$\boldsymbol{\Omega}_i^p = \mathrm{supp}\left(\hat{\boldsymbol{A}}_{\mathrm{res},i}^{\mathrm{aux},p}\right), \tag{129}$$

calculated for each group $i = 1,\dots,N_g$, are the effectively used output of this estimation step. Those supports are assumed to be the same as the supports of the final residual coefficient vectors $\hat{\boldsymbol{A}}^p_{\text{res},1},\dots,\hat{\boldsymbol{A}}^p_{\text{res},N_g}$ whose magnitudes are estimated via linear regression in the next step.

c) *Estimation of Magnitudes of $\hat{\boldsymbol{A}}^p_{\text{res},1},\dots,\hat{\boldsymbol{A}}^p_{\text{res},N_g}$ on Supports Estimated in Step b)*
The bias in magnitude, which is generally introduced by sparse regularization terms, is corrected for by an *a posteriori* re-estimation of the magnitudes of the non-trivial coefficients in $\hat{\boldsymbol{A}}^p_{\text{res,i}}$. This step allows for an incorporation of all three input patches as follows:

$$\begin{aligned}
&\left\{\hat{\boldsymbol{A}}^p_{\text{res},1},\dots,\hat{\boldsymbol{A}}^p_{\text{res},N_g}\right\} = \\
&\min_{\substack{\boldsymbol{A}^p_{\text{res},1},\dots,\boldsymbol{A}^p_{\text{res},N_g} \\ \boldsymbol{A}^p_{\text{res},i}|_{\boldsymbol{\Omega}^{p,c}_i}=\boldsymbol{0} \\ \forall i=1,\dots,N_g}} \Bigg\{ \frac{1}{N^P_{\text{H}}N_g}\sum_{i=1}^{N_g}\frac{1}{n_{c,i}}\left\|\left(\boldsymbol{d}^p_{\text{H},i}\boldsymbol{a}^p_i+\boldsymbol{D}^p_{\text{res,H},i}\boldsymbol{A}^p_{\text{res},i}\right)^T+\boldsymbol{P}^p_i\left(\hat{\boldsymbol{m}}^p_Z\boldsymbol{1}^{1\times N^P_{\text{H}}}-\tilde{\boldsymbol{Z}}^p\right)\right\|^2_F \\
&\quad+\frac{\mu_Y}{N^P_{\text{L}}N_g}\sum_{i=1}^{N_g}\frac{1}{n_{c,i}}\left\|\left(\boldsymbol{d}^p_{\text{L},i}\boldsymbol{a}^p_i+\boldsymbol{D}^p_{\text{res,H},i}\boldsymbol{A}^p_{\text{res},i}\right)^T+\boldsymbol{P}^p_i\left(\boldsymbol{m}^p_Y\boldsymbol{1}^{1\times N^P_{\text{L}}}-\boldsymbol{Y}^p\right)\right\|^2_F \\
&\quad+\frac{\mu_X}{N^P_{\text{H}}N^X_b}\left\|\boldsymbol{R}\left(\widehat{\boldsymbol{P}}^p\left(\widehat{\boldsymbol{d}}^p_{\text{H}}\widehat{\boldsymbol{a}}^p+\widehat{\boldsymbol{D}}^p_{\text{res,H}}\widehat{\boldsymbol{A}}^p_{\text{res}}\right)^T+\hat{\boldsymbol{m}}^p_Z\boldsymbol{1}^{1\times N^P_{\text{H}}}\right)-\boldsymbol{X}^p\right\|^2_F\Bigg\}
\end{aligned} \tag{130}$$

In Eq. (130), the quantity $\boldsymbol{\Omega}^{p,c}_i$ denotes the *complement* of $\boldsymbol{\Omega}^p_i$. Moreover, the condition $\boldsymbol{A}^p_{\text{res},1}|_{\boldsymbol{\Omega}^{p,c}_i}=\boldsymbol{0}$ restricts the variable part of $\boldsymbol{A}^p_{\text{res},i}$ to the support $\boldsymbol{\Omega}^p_i$ while setting all other coefficients to zeros. Finally, $\widehat{\boldsymbol{D}}^p_{\text{res,H}}$ and $\widehat{\boldsymbol{A}}^p_{\text{res}}$ are defined from $\boldsymbol{D}^p_{\text{res,H},i}$ and $\boldsymbol{A}^p_{\text{res},i}$, $i = 1,\dots,N_g$, analogously to how $\widehat{\boldsymbol{D}}^p_{\text{H}}$ and $\widehat{\boldsymbol{A}}^p$ are defined in Eq. (103) and (104).

d) *Construction of Final Jointly Sparse Coefficients $\hat{\boldsymbol{A}}^p_1,\dots,\hat{\boldsymbol{A}}^p_{N_g}$*
For each group of hyperspectral channels, $i = 1,\dots,N_g$, the final group of jointly sparse coefficient vectors is composed of the local coefficient $\hat{\boldsymbol{a}}^p_i$ and the non-local residual coefficients $\hat{\boldsymbol{A}}^p_{\text{res},i}$ as follows:

$$\hat{\boldsymbol{A}}^p_i=\begin{bmatrix}\hat{\boldsymbol{a}}^p_i \\ \hat{\boldsymbol{A}}^p_{\text{res},i}\end{bmatrix}\begin{matrix}\leftarrow \text{local contribution} \\ \leftarrow \text{non-local contribution}\end{matrix} \tag{131}$$

❺ Construction of High-Resolution Hyperspectral Patch $\hat{\boldsymbol{Z}}^p$

With the hyperspectral channel groups $i = 1,\dots,N_g$ and quantities $\widehat{\boldsymbol{P}}^p$ and $\boldsymbol{P}^p_1,\dots,\boldsymbol{P}^p_{N_g}$ calculated in step ❷, the high-resolution dictionaries, $\boldsymbol{D}^p_{\text{H},1},\dots,\boldsymbol{D}^p_{\text{H},N_g}$, calculated in step ❸ and the corresponding groups of jointly sparse coefficient vectors, $\hat{\boldsymbol{A}}^p_1,\dots,\dots,\hat{\boldsymbol{A}}^p_{N_g}$, estimated in step ❹, the band-wise zero-mean hyperspectral patch $\hat{\boldsymbol{Z}}^p_0$ is obtained by following Eqs. (102) through (104).

The final high-resolution hyperspectral patch $\hat{\boldsymbol{Z}}^p$ is then obtained by applying Eq. (101) in which the sample mean values $\hat{\boldsymbol{m}}^p_Z$ of $\hat{\boldsymbol{Z}}^p$ estimated in step ❶ are added to $\hat{\boldsymbol{Z}}^p_0$.

High-resolution Hyperspectral Image Construction
As the last step in the local–non-local processing module, the full hyperspectral image $\boldsymbol{I}_{\hat{Z}}$ is constructed band-by-band by spatially tiling all patches $\hat{\boldsymbol{Z}}^p$, $p = 1,\ldots,N_p$, at their respective positions in the image while averaging at overlapping pixels.

5.2.4 Global Processing Module

In order to ensure the spatial consistency of the fusion product, in particular at patch-boundaries, the global processing module conducts optimization on the full spatial scale simultaneously. Similarly, the full hyperspectral wavelength range is optimized at the same time.

As is depicted in the flowchart of the J-SparseFI-HM algorithm in Fig. 51 on page 106, the input quantities of this module are similar to those of the local–non-local processing module. In particular, the hyper- and multispectral input data $\boldsymbol{I}_Y$ and $\boldsymbol{I}_X$, the output high-resolution data $\boldsymbol{I}_{\tilde{Z}}$ of the local–non-local module as well as information about the system's SRFs and PSFs are required. The SRFs and PSFs are, again, represented by the matrices $\boldsymbol{R} \in \mathbb{R}^{N_b^X \times N_b^Y}$ and $\boldsymbol{S} \in \mathbb{R}^{N_p^X \times N_p^Y}$, respectively.

The idea is to solve a modified version of the originally ill-posed linear system given by the observation model in Eqs. (99) and (100) on the full image scale. The problem is rendered well-posed by incorporating information about the previously generated intermediate fusion product $\boldsymbol{I}_{\tilde{Z}}$ and performing optimization on a linear spectral subspace.

Spectral Subspace Transformation The rationale for optimizing on a spectral subspace is two-fold: On the one hand, it reduces computational complexity. On the other hand – and this is the primary driving factor –, the denoising capacity of subspace projections calculated via, e.g., *principal component analysis* (PCA) or *singular value decomposition* (SVD) are well-suited for compensating for the system and data noise represented by the terms $\boldsymbol{E}_X$ and $\boldsymbol{E}_Y$ in the observation model. In particular, the new estimate $\boldsymbol{I}_{\hat{Z}}$ is found via

$$\boldsymbol{I}_{\hat{Z}} = \boldsymbol{W}^T \boldsymbol{J}_{\hat{Z}} \tag{132}$$

where

- $\boldsymbol{W} \in \mathbb{R}^{N_{b,\mathrm{sub}}^Y \times N_b^Y}$ is a linear transformation matrix, which projects full hyperspectral profiles of dimension N_b^Y onto the spectral subspace spanned by the rows of $\boldsymbol{W}$; As will be explained below, $\boldsymbol{W}$ can be directly derived from the hyperspectral input image $\boldsymbol{I}_Y$;
- $\boldsymbol{J}_{\hat{Z}} \in \mathbb{R}^{N_{b,\mathrm{sub}}^Y \times N_p^X}$ contains the corresponding representation coefficients to be estimated.

As was elaborated on in Section 2.1.6, there are various ways of calculating a spectral subspace transformation matrix and the intrinsic subspace dimension. In the work towards this thesis, both the *singular value decomposition* (SVD) and the popular spectral unmixing-based *vertex component analysis* (VCA) [Nascimento and Dias, 2005b] have been tested and compared. It was found that directly deriving $\boldsymbol{W}$ from the SVD representation of $\boldsymbol{I}_Y$ yields similar yet slightly better representation results than those achieved using VCA. Therefore, VCA is no further discussed here. The SVD-based transformation matrix is obtained as follows:

$$\boldsymbol{W} = \boldsymbol{U}_{\mathrm{sub}}\,, \tag{133}$$

where $\boldsymbol{U}_{\text{sub}}$ is composed of the leading $N^Y_{b,\text{sub}}$ left-singular vectors of $\boldsymbol{I}_Y$. That is, the columns of $\boldsymbol{U}_{\text{sub}}$ equal the first $N^Y_{b,\text{sub}}$ columns of $\boldsymbol{U} \in \mathbb{R}^{N^Y_b \times N^Y_b}$ which, in turn, is the unitary left-most matrix in the SVD of the input hyperspectral image $\boldsymbol{I}_Y$, i.e.

$$\boldsymbol{I}_Y = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T. \tag{134}$$

In Eq. (134), $\boldsymbol{\Sigma}$ is a rectangular diagonal matrix containing the singular values of $\boldsymbol{I}_Y$ in – what is assumed here to be – non-increasing order and $\boldsymbol{V}$ is the other – momentarily irrelevant – unitary matrix constituting the spatial information in $\boldsymbol{I}_Y$.

Since PCA is often used in the literature in similar contexts, it is worth to note that in the particular problem at hand, where only the spectral projection matrix $\boldsymbol{U}$, i.e. $\boldsymbol{W}$, is used and remains fix while $\boldsymbol{J}_{\hat{Z}}$ is optimized for in an unconstrained manner in Eq. (138) further below, PCA and SVD can be used *interchangeably* for spectral subspace projection without changing the product $\boldsymbol{I}_{\hat{Z}}$. This is because both $\boldsymbol{U}$ and the spectral transformation matrix obtained through PCA constitute eigenvectors of $\boldsymbol{I}_Y\boldsymbol{I}_Y^T$ as columns. The possible scalar differences between corresponding columns are automatically compensated for through $\boldsymbol{J}_{\hat{Z}}$ in Eq. (138).

As for the determination of the intrinsic spectral subspace dimension $N^Y_{b,\text{sub}}$, there are a number of methods developed specifically for identifying the subspace of hyperspectral data [Nascimento and Dias, 2005a; Bioucas-Dias and Nascimento, 2008; Rasti et al., 2015b]. In principal, any of these techniques can be used for this purpose. In this work, however, it has been found that $N^Y_{b,\text{sub}}$ matches most reliably with the feature band, beyond which the sorted data variance "flattens" (see Fig. 24 (a) on page 35). This feature band index can be well identified directly from the data variance. In Section 5.4, a sample analysis of the impact of $N^Y_{b,\text{sub}}$ on the fusion performance will reveal that peak performance can indeed be expected near the flattening point of the variance curve.

Optimization Problem In order to incorporate the assumption of multi- and hyperspectral input data being corrupted by colored – i.e. band-dependent – but spatially comparatively lowly dependent noise, each row in the error matrices $\boldsymbol{E}_X$ and $\boldsymbol{E}_Y$ is assumed to be individually describable by an independent identically distributed (i.i.d.) random process. As was proposed first by Hardie et al. [2004] in the context of hyperspectral-panchromatic data fusion and adopted by Wei et al. [2015a] for hyperspectral-multispectral data fusion more recently, fusion problems that are modeled using the above-described sensor observation relationship (see Eqs. (99) and (100)) can be treated as a maximum *a posteriori* (MAP) estimation process. In the resulting minimization problem, the colored noise assumption is naturally incorporated through the multi- and hyperspectral noise covariance matrices $\boldsymbol{\Sigma}_X \in \mathbb{R}^{N^X_b \times N^X_b}$ and $\boldsymbol{\Sigma}_Y \in \mathbb{R}^{N^Y_b \times N^Y_b}$, respectively. If the noise components in different channels are assumed to be mutually uncorrelated, i.e. $\boldsymbol{\Sigma}_X$ and $\boldsymbol{\Sigma}_Y$ are diagonal matrices, contributions of the channel-specific residual errors to the cost function value are weighed by the corresponding noise variances that are stored on the diagonals of $\boldsymbol{\Sigma}_X$ and $\boldsymbol{\Sigma}_Y$.

This approach is adopted here and extended in a way that allows accounting for the potentially incomplete estimate $\boldsymbol{I}_{\tilde{Z}}$ in addition to the original input data $\boldsymbol{I}_X$ and $\boldsymbol{I}_Y$. In particular, the extended system model is described by

$$\boldsymbol{I}_X = \boldsymbol{R}\boldsymbol{I}_Z + \boldsymbol{E}_X \tag{135}$$
$$\boldsymbol{I}_Y = \boldsymbol{I}_Z\boldsymbol{S} + \boldsymbol{E}_Y \tag{136}$$
$$\boldsymbol{I}_{\tilde{Z}} = \boldsymbol{I}_Z + \boldsymbol{E}_{\tilde{Z}}, \tag{137}$$

where $\boldsymbol{E}_{\tilde{Z}}$ denotes the colored, spatially independent error between the target fusion product $\boldsymbol{I}_Z$ and the previous estimate $\boldsymbol{I}_{\tilde{Z}}$. Furthermore, the covariance matrices corresponding to $\boldsymbol{E}_X$, $\boldsymbol{E}_Y$ and $\boldsymbol{E}_{\tilde{Z}}$ are assumed to be describable as $\Sigma_X = \mathrm{diag}(\sigma^2_{X,1},\ldots,\sigma^2_{X,N_b^X})$, $\Sigma_Y = \mathrm{diag}(\sigma^2_{Y,1},\ldots,\sigma^2_{Y,N_b^Y})$ and $\Sigma_{\tilde{Z}} = \mathrm{diag}(\sigma^2_{\tilde{Z},1},\ldots,\sigma^2_{\tilde{Z},N_b^Y})$, where $\sigma_{X,k}$, $\sigma_{Y,l}$ and $\sigma_{\tilde{Z},l}$ denote the standard deviations of $\boldsymbol{E}_X$, $\boldsymbol{E}_Y$ and $\boldsymbol{E}_{\tilde{Z}}$, respectively, in bands $k = 1,\ldots,N_b^X$ and $l = 1,\ldots,N_b^Y$.

With $\boldsymbol{A}^{-1/2}$ denoting the element-wise nested invert and square-root operator applied to any real-valued matrix $\boldsymbol{A}$, the resulting optimization problem can be written as follows:

$$\boldsymbol{J}_{\hat{Z}} = \min_{\boldsymbol{J}_Z}\left\{\mu'_Z\left\|\Sigma_{\tilde{Z}}^{-1/2}\left(\boldsymbol{W}^T\boldsymbol{J}_Z - \boldsymbol{I}_{\tilde{Z}}\right)\right\|_F^2 + \left\|\Sigma_X^{-1/2}\left(\boldsymbol{R}\boldsymbol{W}^T\boldsymbol{J}_Z - \boldsymbol{I}_X\right)\right\|_F^2 + \left\|\Sigma_Y^{-1/2}\left(\boldsymbol{W}^T\boldsymbol{J}_Z\boldsymbol{S} - \boldsymbol{I}_Y\right)\right\|_F^2\right\} \quad (138)$$

The parameter $\mu'_{\tilde{Z}}$ is used to regulate the impact of the previous estimate $\boldsymbol{I}_{\tilde{Z}}$ to the fusion result relative to the impact of the original input images $\boldsymbol{I}_X$ and $\boldsymbol{I}_Y$. Note that, instead of using a single regularization parameter $\mu'_{\tilde{Z}}$, Eq. (138) can be principally equipped with two parameters μ'_X and μ'_Y before the corresponding center- and right-most fidelity terms. However, a large number and variety of tests during the research towards this thesis have revealed that the potential fusion quality improvement achievable through those two parameters does not outweigh the additional complexity brought by the necessity of tuning two parameters instead of one.

5.2.5 Data-driven Determination of System Characteristics

Up to this point, the relative spectral and spatial sensor responses stored in the matrices $\boldsymbol{R}$ and $\boldsymbol{S}$ as well as the noise covariance matrices Σ_X, Σ_Y and $\Sigma_{\tilde{Z}}$ have been assumed given. The task of acquiring this information is, sometimes, treated separately from the data fusion task in the literature [Yokoya et al., 2012, 2013; Lanaras et al., 2015, 2016]. Most often, however, this topic is not addressed at all.

It is important to concede that accurate multi-sensor data fusion requires more than the possession and usage of individual sensor information that was acquired during laboratory measurements conducted prior to the operational phase. On the one hand, actual sensor characteristics differ from the original sensor specifications due to aspects such as aging of the optomechanical and electronic instrument components and the substantial difference of the orbital/aerial environment to the laboratory [Wang et al., 2010]. On the other hand, a view to Eqs. (135) and (136) suggests that the matrices $\boldsymbol{R}$ and $\boldsymbol{S}$ may, effectively, carry more information than the mere re-sampled SRFs and PSFs. They rather link the two input data sets to the fusion product. In particular, they potentially capture non-stationary, data acquisition-specific details about the observations, such as differences in illumination and atmospheric conditions, and different viewing angles. Furthermore, the unavoidable imperfection in image co-registration of the input images negatively affects the quality of fusion products [Yokoya, Grohnfeldt, and Chanussot, 2017]. These factors may be at least partially compensated by adjusting $\boldsymbol{R}$ and $\boldsymbol{S}$ to the data at hand.

Considering the observation model given by Eqs. (135) and (136), a natural starting point for estimating the system's spectral and spatial characteristics in a data-driven manner is the following relationship between the input images $\boldsymbol{I}_X$ and $\boldsymbol{I}_Y$:

$$\boldsymbol{R}\boldsymbol{I}_Y = \boldsymbol{I}_X\boldsymbol{S} + \boldsymbol{E}_{XY}\,. \quad (139)$$

In Eq. (139), the noise terms $\boldsymbol{E}_X$ and $\boldsymbol{E}_Y$ are implicitly incorporated in the relative error $\boldsymbol{E}_{XY}$. The idea is to minimize $\boldsymbol{E}_{XY}$ and, therefore, $\boldsymbol{E}_X$ and $\boldsymbol{E}_Y$ by optimizing for $\boldsymbol{R}$ and $\boldsymbol{S}$ via linear regression by solving a possibly constrained or regularized version of the following problem:

$$\min_{\boldsymbol{R},\boldsymbol{S}} \|\boldsymbol{R}\boldsymbol{I}_Y - \boldsymbol{I}_X\boldsymbol{S}\|_F^2 \tag{140}$$

This approach has been followed in the recent literature on this topic [Yokoya et al., 2013; Simões et al., 2015; Lanaras et al., 2016]. The main differences between the existing methods are the assumptions and constraints on the target variables. One idea, which is implemented in all of the methods proposed in the above-mentioned publications, is that $\boldsymbol{R}$ and $\boldsymbol{S}$ are optimized for neither simultaneously nor in an alternating manner. Instead, Simões et al. [2015] optimize for $\boldsymbol{R}$ and $\boldsymbol{S}$ sequentially in that order, subject to quadratic total variation (TV) penalty terms while *not* enforcing non-negativity. In contrast, Yokoya et al. [2013] and Lanaras et al. [2016] both propose to optimize for $\boldsymbol{S}$ first and then for $\boldsymbol{R}$ while constraining $\boldsymbol{R}$ to contain only non-negative elements. Moreover, both Yokoya et al. [2013] and Lanaras et al. [2016] propose estimating $\boldsymbol{S}$ using the original relative SRFs provided by the sensor's manufacturer.

Point Spread Functions Stored in $\boldsymbol{S}$ Each column in $\boldsymbol{S}$ contains one spatial blurring filter corresponding to one low-resolution hyperspectral pixel. Lanaras et al. [2016] constrain this filter to be separable and symmetric. In [Yokoya et al., 2013], the discrete spatial blurring function is assumed to be a 2-D Gaussian filter centered at the center of the hyperspectral pixel with a pre-defined width but variable variance. In this work, the latter approach is followed, but with a fixed variance and sum-to-one contraint to make the filter's full width at half maximum match with the dimension of the low-resolution pixel. This way, the discrete Gaussian function approximates the net PSF model described in Eq. (13) on page 15 which is commonly used in image fusion [Yokoya et al., 2012; Zhu et al., 2016] and has consistently led to realistic results in the numerous tests conducted during the research towards this thesis.

Relative Spectral Responses Stored in $\boldsymbol{R}$ Given the matrix $\boldsymbol{S}$, a cost function equal or similar to $\|\boldsymbol{R}\boldsymbol{I}_Y - \boldsymbol{I}_X\boldsymbol{S}\|_F^2$ (see Eq. (140)) remains to be minimized for $\boldsymbol{R}$. Note that the rows $\boldsymbol{r}_l$, $l = 1,\ldots,N_b^X$, of $\boldsymbol{R}$, each of which corresponds to one of N_b^X multispectral channels, can be optimized for individually. Simões et al. [2015] regularize this problem with a quadratic total variation penalty on $\boldsymbol{r}_i$, which, according to Lanaras et al. [2016], tend to over-smooth the result and is particularly unsuitable for the approximation of rather rectangular-shaped SRFs such as those of Landsat-8. For such SRFs, Lanaras et al. [2016] propose changing the quadratic regularization term to an l_1-norm-based TV-regularization while adding a non-negativity constraint on $\boldsymbol{r}_i$. Also Yokoya et al. [2013] report that the physically meaningful non-negativity constraint generally leads to fusion quality improvements. In contrast to the other proposed methods, Yokoya et al. [2013] tightly constrain the rows of $\boldsymbol{R}$ to deviate from the known SRFs only up to a given tolerance $\epsilon > 0$.

In this work, the error $\boldsymbol{E}_{XY}$ is further minimized by intentionally keeping the degrees of freedom high in the estimation procedure for $\boldsymbol{R}$. The only constraint on $\boldsymbol{R}$ that has been found to consistently yield overall fusion quality improvements in numerous tests conducted on this topic, is, indeed, the non-negativity of $\boldsymbol{R}$. Moreover, factors such as different illumination conditions during the acquisition of $\boldsymbol{I}_X$ and $\boldsymbol{I}_Y$ or other causes for the common differences in intensity between the two input images are addressed by optimizing for a translative intensity difference $\boldsymbol{t} \in \mathbb{R}^{N_b^X}$ between the multispectral channels in $\boldsymbol{I}_X$ and weighted sums of hyperspectral channels in $\boldsymbol{I}_Y$ alongside $\boldsymbol{R}$. The resulting optimization problem, that is used

in this work, is given by

$$\min_{\boldsymbol{R},\boldsymbol{t}} \|\boldsymbol{R}\boldsymbol{I}_Y - (\boldsymbol{I}_X + \boldsymbol{t}\mathbf{1}^{1\times N_p^X})\boldsymbol{S}\|_F^2 \quad \text{s.t.} \quad \boldsymbol{R} \geq \mathbf{0}^{N_b^X \times N_b^Y}, \tag{141}$$

where $\mathbf{0}^{n\times m}$ denotes a matrix of dimension $n \times m$ that contains only zeros and the operator "$\geq$" is meant to be applied element-wise. Subsequently to solving Eq. (141), the estimated intensity shift $\boldsymbol{t}$ is added to the input image $\boldsymbol{I}_X$ prior to estimating $\boldsymbol{I}_Z$, i.e. the actual fusion process. That is, when applying the above-described procedure for estimating $\boldsymbol{R}$, the multispectral image denoted by $\boldsymbol{I}_X$ in Sections 5.2.1 through 5.2.4 actually refers to the spectrally translated image $\boldsymbol{I}_{X,\text{translated}} = \boldsymbol{I}_X + \boldsymbol{t}\mathbf{1}^{1\times N_p^X}$. However, this technical detail is only mentioned here for the sake of simplicity.

Noise Variances Stored in Σ_X, Σ_Y, and $\Sigma_{\tilde{Z}}$ In order to solve Eq. (138), the noise covariance matrices Σ_X, Σ_Y, and $\Sigma_{\tilde{Z}}$ need to be known or estimated from the data.

As it was mentioned in Section 5.2.4, in this work, the noise components in different bands are assumed to be spectrally uncorrelated. This assumption is common in the field of multi- and hyperspectral data modelling [Acito et al., 2011], [Landgrebe, 2005, ch. 9]. It implicates that the spectral covariance matrix contains non-zero elements only along the diagonal. Hence, the problem of estimating the covariance matrix simplifies to the problem of estimating the noise variance in each spectral channel. This assumption also implicates that, in each of the three terms in Eq. (138), the contribution of the residual error in a specific band to the cost function is weighed solely by the inverse of the noise variance.

Regarding the estimation of the noise level of an image, a large variety of methods have been proposed to solve this problem in the past. Single-band noise estimation algorithms often rely – in one way or another – on finding homogeneous areas in which the deviation from the mean is associated with noise [Gao et al., 2008; Liu et al., 2013]. This approach is particularly effective for data of high spatial resolution, which is why it is used, in this work, to estimate Σ_X, i.e. the noise variances of $\boldsymbol{I}_X$. More specifically, the method recently proposed by Liu et al. [2013] is used, which has empirically shown to reliably produce reasonable results for moderate- to high-spatial resolution multispectral data. A corresponding experiment substantiating this statement is presented in Section 5.4.1. The method is convenient to use, as it automatically detects low-rank patches – representing homogeneous areas – based on gradients and local statistics. The noise level is then estimated from those patches using PCA.

Considering the low spatial resolution of space-based hyperspectral sensors, the principle of relying on finding homogeneous areas in the image tends to lead to an over-estimation of the noise level if no spectral information is additionally incorporated in the estimation procedure. As will be experimentally demonstrated in Section 5.4.1, this problem is particularly prominent when such a method is applied to hyperspectral data that was acquired over highly heterogeneous urban ground where homogeneous regions rarely occur at lower spatial resolutions. In this case, the noise level should be estimated primarily based on the rich *spectral* information in hyperspectral imagery. One of the most established methods, which has been originally developed to reliably estimation the spectral noise variance in AVIRIS hyperspectral data but has been serving as a benchmark method for general hyperspectral noise estimation since then [Acito et al., 2011; Gao et al., 2013], is the linear regression-based spectral and spatial de-correlation (SSDC) method by Roger and Arnold [1996]. Due to its effectiveness and simplicity, SSDC is used in this work to estimate the variances of the hyperspectral data sets, i.e. Σ_Y, and $\Sigma_{\tilde{Z}}$.

5.3 Experimental Setup

This section provides detailed information about the experiments conducted in this chapter. A large variety of test scenarios including various sensor combinations featuring different spectral and spatial characteristics has been compiled in order to test the capacity and limits of the proposed fusion methodology. The following Subsections 5.3.1 through 5.3.5 introduce the underlying data sets, give information about specific simulation procedures, briefly discuss the fusion quality assessment criteria used in the chapter, introduce the fusion algorithms used for comparison, and provide information about the implementation, libraries, compilers and computer platform used for processing.

5.3.1 Data sets

In order to maximize transparency and comparability to the state of the art throughout the entire experimental part, the data sets used for testing are the same as those prepared for and used in the review article on hyperspectral and multispectral data fusion that was published recently in IEEE Geoscience and Remote Sensing Magazine [Yokoya, Grohnfeldt, and Chanussot, 2017].

Tab. 6 summarized the main specifications of all the data sets used in this work. Moreover, corresponding RGB true-color compositions are depicted in Fig. 55.

These eight data sets feature different geographical and sensor characteristics. They were selected and prepared to thoroughly test the versatility and generalizability of the algorithms that were proposed recently for the fusion of hyper- and multispectral data. The observed scenes contain a diversity of materials and structures (e.g. different types of vegetation, minerals and urban structures) which implicate potential applications for high-resolution hyperspectral data products.

The following paragraphs briefly describe each data set.

CASI – University of Houston This data set was used in the 2013 IEEE Geosciences and Remote Sensing Society Data Fusion Contest [Debes et al., 2014]. The image was acquired by an ITRES Compact Airborne Spectographic Imager (CASI)-1500 sensor over the campus of the University of Houston and its surrounding urban areas. The instrument measures signals in 144 distinct spectral channels covering the wavelength range from 0.364 to 1.046 μm. The original data scene has a spatial dimension of $349 \times 1{,}905$ pixels with a GSD of 2.5 m. For our recent comparative study [Yokoya, Grohnfeldt, and Chanussot, 2017], a representative sub-area of 320×540 pixels was selected. An RGB true-color composition of this area is shown in Fig. 55 (a).

Hyperspec – Chikusei The airborne hyperspectral data set was acquired by Headwall's Hyperspec-VNIR-C imaging sensor over the city of Chikusei, Ibaraki Prefecture, Japan, on 29 July 2014. The data set comprises 128 bands in the spectral range from 0.363 to 1.018 μm. The original scene consists of $2{,}517 \times 2{,}335$ pixels at a GSD of 2.5 m. The selected spatial subset of size 540×420 pixels shows a suburban area dominated by agricultural country (see Fig. 55 (b)). Detailed information about this data set including data acquisition and processing aspects was made available to the scientific community recently [Yokoya and Iwasaki, 2016].

ROSIS-3 – University of Pavia The Reflective Optics System Imaging Spectrometer (ROSIS-3) acquired this data set over the University of Pavia, Italy, in 2003. It consists of

Data set index	1	2	3	4	5	6	7	8
Site	University of Houston	Chikusei	University of Pavia	Washington D.C. Mall	Indian Pines	Cuprite	Moffett Field	Rodalquilar
Original sensor	CASI	Hyperspec	ROSIS-3	HYDICE	AVIRIS	AVIRIS	AVIRIS	HyMap (EnMAP)
Year	2012	2014	2003	1995	1992	1995	1997	2003
Data type	Radiance	Reflectance	Reflectance	DN	Radiance	Reflectance	Radiance	Reflectance
Spectral range [μm]	0.36–1.05	0.36–1.02	0.43–0.84	0.4–2.5	0.4–2.5	0.4–2.5	0.4–2.5	0.4–2.5
Number of bands	144	128	115	210	224	224	224	126 (242)
Used bands	144	128	103	191	192	185	182	167
GSD	2.5	2.5	1.3	2.5	20	20	17	4 (10)
Spatial dimension (number of pixels)	320×540	540×420	560×320	420×300	360×360	420×360	360×360	261×867
Multispectral sensor	Sentinel-2	WV-2	QB	QB	WV-3	WV-3	QB	Sentinel-2
GSD ratio	5	6	8	4	4	5	4	3
SNR [dB]	35	35	35	35	35	35	35	n.a.

Table 6. Specifications of the data sets used in the experimental sections.

610×340 pixels with a GSD of 1.3 m. After removing 12 bands of very low SNR, a total of 103 bands covering the spectral range from 0.430 to 0.838 μm remain to be used in the experiments. The image is available as reflectance data, and a 560×320-pixel-size image was selected for the experiments (see Fig. 55 (c)).

HYDICE – Washington, D.C., The National Mall This image was taken with the Hyperspectral Digital Imagery Collection Experiment (HYDICE) sensor [Mitchell, 1995] over the National Mall in Washington, D.C., in 1995. The HYDICE sensor acquired 210 bands between 0.4 and 2.5 μm, and the full image consists of $1{,}280\times307$ pixels at a GSD of 2.5 m. A 420×300-pixel-size subimage was selected (see Fig.55 (d)) and 191 bands are used after removing water absorption bands in the 0.9–1.4 μm region.

AVIRIS – Indian Pines This hyperspectral image was acquired by the Airborne Visible/InfraRed Imaging Spectrometer (AVIRIS) sensor [Vane et al., 1993] over the Indian Pines test site in northwestern Indiana in 1992 [Baumgardner et al., 2015]. The AVIRIS sensor acquires 224 spectral bands in the wavelength range from 0.4 to 2.5 μm with a full width at half maximum (FWHM) of 10 nm. The image consists of 512×614 pixels at a GSD of 20 m. A 360×360-pixel-size subimage (see Fig. 55 (e)) was selected and 192 bands are used after removing bands of strong water vapor absorption and low SNRs.

AVIRIS – Cuprite The AVIRIS sensor acquired this data set in 1995 over the Cuprite mining district in Nevada. Like the Indian Pines and the Moffett Field AVIRIS data sets used in this work, the full data is publicly available on `http://aviris.jpl.nasa.gov/data/free_data.html`. The full Cuprite data set comprises five reflectance images. The one that is used in this work is stored on the aforementioned website under the file name *f970619t01p02_r02_sc03.a.rfl*. The scene consists of 512×614 pixels at a GSD of 20 m. A spatial subarea of size 420×360 pixels was selected (see Fig. 55(f)) and 185 bands are used after removing noisy bands.

AVIRIS – Moffett Field This data set was captured in 1997 by the AVIRIS sensor over the Moffett Field in California. The original calibrated radiance image consists of $1{,}923\times753$ pixels and has a GSD of 17 m. A 360×360-pixel-size subimage was selected (see Fig. 55(g)) and 182 bands are used after removing water absorption bands.

HyMap – Rodalquilar The HyMap image was captured over Rodaquilar, Spain, in June 2003 [Bedini et al., 2009]. The HyMap sensor collected 126 bands in the 0.4–2.5-μm wave-

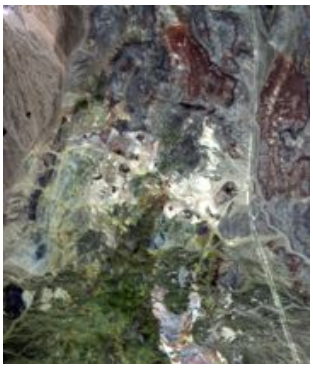
(a) CASI: *University of Houston*

(b) Hyperspec: *Chikusei*

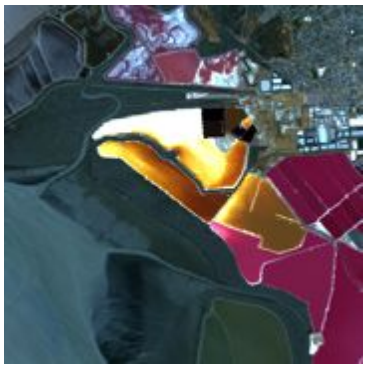
(c) ROSIS-3: *Pavia University*

(d) HYDICE: *Washington D.C. Mall*

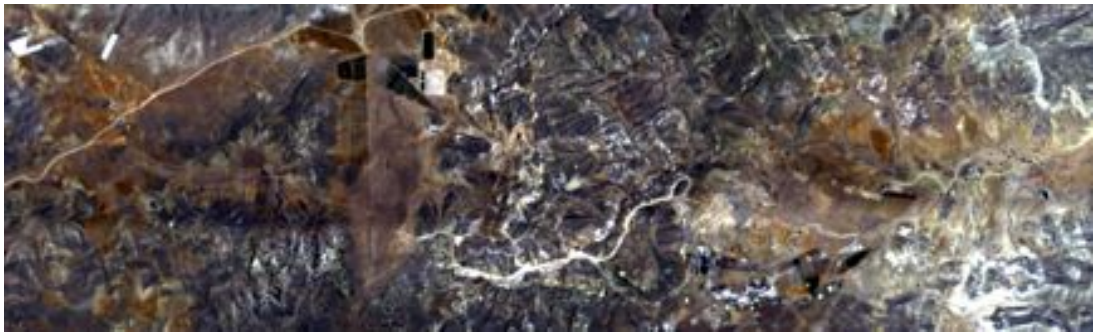
(e) AVIRIS: *Indian Pines*



(f) AVIRIS: *Cuprite*



(g) AVIRIS: *Moffett Field*



(h) HyMap (simulated EnMAP): *Rodalquilar*

Fig. 55. True-color composite images of the original hyperspectral data sets used in the experimental sections. Specifications of each data set can be found in Tab. 6.

length range. The scene covers a gold-mining area in the Sierra del Cabo de Gata (Cabo de Gata National Park) with a GSD of 4 m. To use this data set for a realistic simulation scenario of EnMAP and Sentinel-2 data fusion, the HyMap image was spatially downgraded to a 10-m GSD and spectrally resampled to 242 bands using the SRFs of the EnMAP instrument. As will be further explained in the following section, the simulation procedure used for this test data set differs from the one used for the other seven data sets. For this data set, sophisticated end-to-end-simulation procedures were used which were specifically developed for realistically simulating EnMAP [Segl et al., 2012] and Sentinel-2 [Segl et al., 2015] data. A brief descriptions of those simulation methodologies is given below in the paragraph titled "End-to-End Simulation". A representative sub-image of size 261×867 pixels was selected to be used in the experiments (see Fig. 55 (h)) and 167 bands remain after removing water absorption and other very low SNR bands.

5.3.2 Simulation Procedures

Similarly to the assessment of the pan-sharpening products in Chapter 4, an accurate quality evaluation of the hyperspectral-multispectral data fusion products can be performed best within a simulation study [Simões et al., 2015; Yokoya et al., 2017]. The quality-assessment procedures used in this chapter mainly comprise three steps: 1) simulate input hyperspectral-multispectral images from a (denoised) reference high-resolution hyperspectral image, 2) generate a high-resolution hyperspectral image by fusing the two input images, and 3) compare the fusion product to the reference image.

The eight hyperspectral data sets described in the previous section are used as reference after denoising. In the literature, the original high-resolution hyperspectral data is often used directly as references to conduct quality assessment; however, in many cases the original data set is, itself, corrupted by noise. Obviously, noisy data is not well-suited as reference. In order to increase the SNRs of the reference data and, therefore, improve the reliability of the quantitative evaluation results, the established hyperspectral data denoising method proposed by Roger [1996] has been applied to the original data sets.

For the simulation of the two input images, two strategies have been followed: 1) A "standard" simulation procedure that considers spectral simulation, spatial simulation, and noise simulation and 2) a more sophisticated end-to-end simulation procedure which takes into account the entire image acquisition and processing chains of EnMAP [Segl et al., 2012] and Sentinel-2 [Segl et al., 2015], starting from raw data. The first strategy is used for generating input data for the fusion of data sets 1 through 7, and the second strategy is used to generate input data corresponding to the HyMap – *Rodalquilar* data set number 8.

Standard Simulation The first simulation procedure is referred to as "standard" procedure here, because it is commonly used for performance evaluation of optical multiresolution data fusion algorithms in the recent literature [Loncan et al., 2015; Simões et al., 2015; Wei et al., 2015c; Zhu et al., 2016; Yokoya et al., 2017]. Except for the denoising component, which was added newly in [Yokoya et al., 2017], this simulation approach is described in Section 2.2.3 and was used in the experimental section of the previous chapter on J-SparseFI for pan-sharpening.

A flow diagram of the evaluation methodology using this standard simulation procedure including denoising is shown in Fig. 56. Simultaneous spectral filtering and down-sampling is performed to generate the multispectral image. In particular, the denoised reference image is spectrally degraded using the SRFs of the desired multispectral sensor as filters. In order to assess and compare the versatility of the proposed data fusion methodology relative to

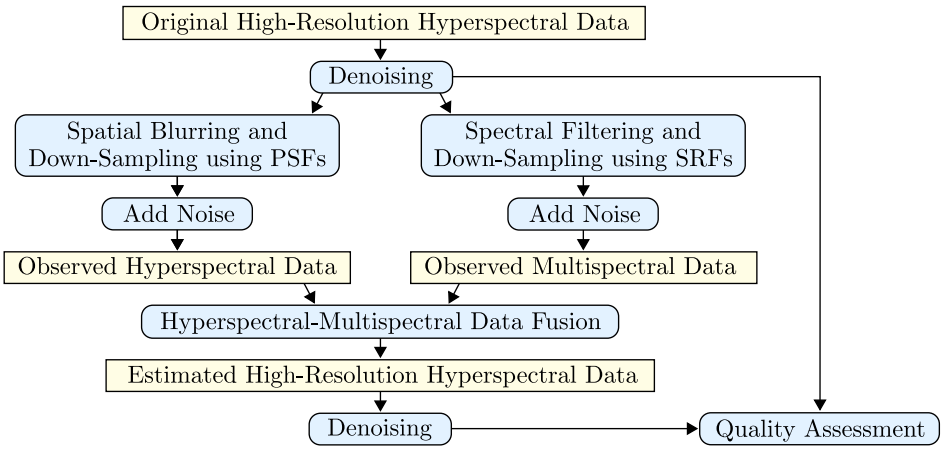


Fig. 56. Flow diagram of the evaluation methodology using the standard simulation procedure.

the state of the art, a variety of multispectral sensors featuring different spectral responses have been selected for simulation, i.e., Sentinel-2 10-m GSD VNIR bands for data set 1; WorldView-2 for data set 2; QuickBird for data sets 3, 4 and 7; and WorldView-3 for data sets 5 and 6 (see Tab. 6).

Fig. 57 shows the relative SRFs of both the hyperspectral and the multispectral imager corresponding to each data set. Noticeable, for data sets 2, 3, 5, and 6, the SRFs of the multispectral sensors cover most of the spectral range of the corresponding hyperspectral imagers. In contrast, there is no high-resolution multispectral information available in the SWIR range for data sets 3, 7, and 8. The latter case is more challenging and was included in the experiments to investigate the impact of the SRF overlap between hyperspectral and multispectral imagers on the quality of the fusion products.

Spatial simulation was performed to generate the low-resolution hyperspectral data using an isotropic Gaussian PSF with a FWHM equalling the GSDs of the low-resolution hyperspectral sensor. Four different GSD ratios, i.e., 4, 5, 6, and 8, are used for spatial simulation in the standard simulation in addition to the GSD of 3 in the end-to-end simulation (see Tab. 6). This variety of resolution ratios increases the diversity of realistic spaceborne sensor combinations and is useful to analyse potential limits and advantages of J-SparseFI-HM relative to the other fusion methods under comparison. Subsequently to the independent spectral and spatial simulation components, colored – i.e. band-dependent – Gaussian noise is added to the simulated hyperspectral and multispectral input images. Realistic noise conditions are attained by simulating SNRs of 35 dB throughout the data sets. The full list of sensor and simulation specifications is given in Tab. 6.

End-to-End Simulation For data set 8, EnMAP and Sentinel-2 L2a (orthorectified surface reflectance data) products were simulated using the *EnMAP end-to-end Simulation* (EeteS) [Segl et al., 2012] and *Sentinel-2 end-to-end Simulation* (S2eteS) [Segl et al., 2015] procedures, respectively. These tools comprise forward and backward simulators that simulate the data-acquisition process as well as the calibration and processing chains, respectively, from spatially and spectrally oversampled data to the final EnMAP and Sentinel-2 products. The forward simulator consists of four independent atmospheric, spatial, spectral, and radiometric modules. The spatial and spectral modules include resampling in the spatial and spectral domains using the sensor-specific PSFs and SRFs, respectively. The radiometric module transforms the at-sensor radiance to digital numbers (DN) by simulating instrumental noise and calibration coefficients. [Yokoya, Grohnfeldt, and Chanussot, 2017].

The backward simulator consists of calibration modules such as non-linearity, dark current,

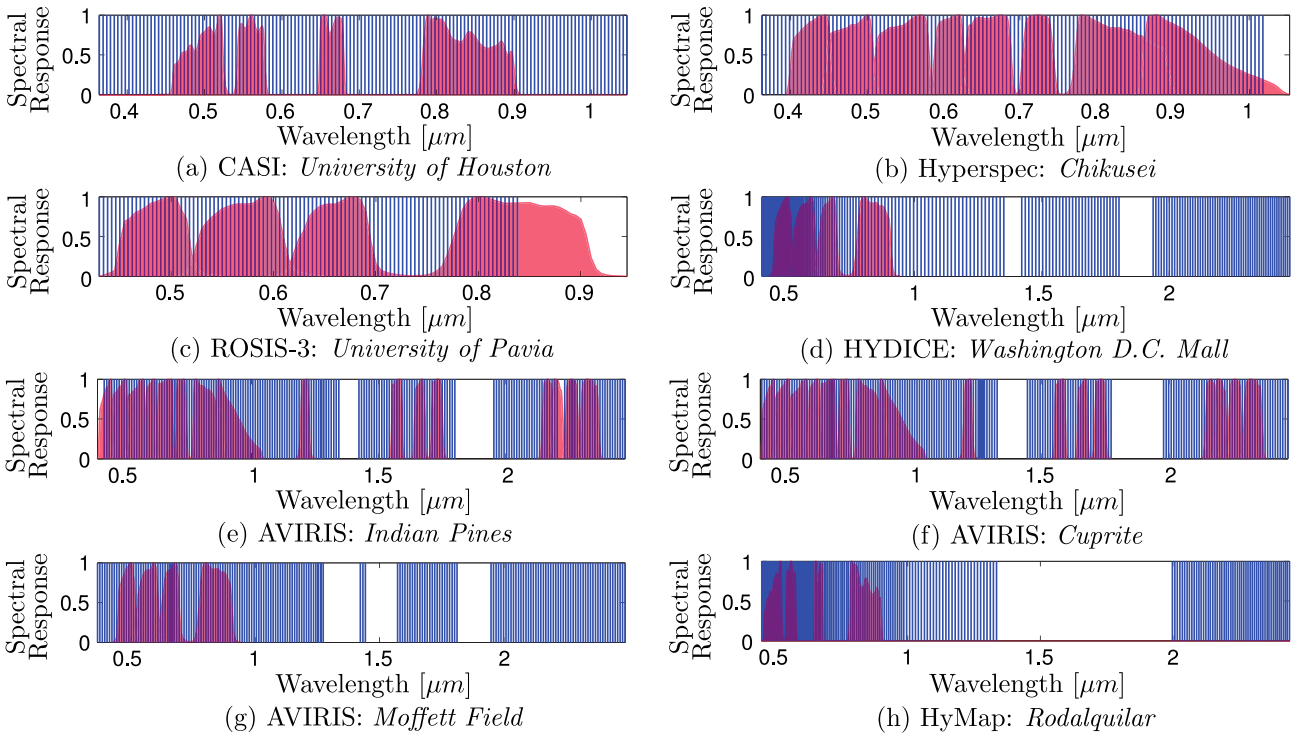


Fig. 57. Center wavelengths of hyperspectral SRFs (blue bars) and relative curves of multispectral SRFs (red curves and shaded areas) corresponding to the data sets used in the experimental sections. Specifications of each data set are summarized in Tab. 6.

and absolute radiometric calibration and preprocessing modules such as radiometric calibration and atmospheric correction. Compared to the standard simulation, the end-to-end simulation can generate more realistic data sets that include errors such as sensor-specific noise and residual errors of atmospheric correction. Sentinel-2 VNIR images with a GSD of 10 m (bands 2, 3, 4, and 8) are used as the multispectral data, even though the 20-m GSD SWIR images could potentially be used in addition for enhancing the EnMAP image, which is of a 30-m GSD.

5.3.3 Fusion Assessment Criteria

In order to assess the quality of the fusion products and analyse the performance of the proposed J-SparseFI-HM algorithm including its sensitivity to variations in system parameters, the following four complementary and widely used quality measures are used for quantitative evaluation: 1) peak SNR (PSNR), 2) spectral angle mapper (SAM) 3), erreur relative globale adimensionnelle de synthèse (ERGAS), and 4) $Q2^n$. Definitions of each of these metrics as well as references to the original literature are provided in Section 2.2.4. ERGAS and $Q2^n$ are designed to simultaneously incorporate both spectral and spatial quality assessment. Hence, they each produce a single number which indicates the overall quality of the fusion product relative to a multiband – here hyperspectral – reference image. In contrast, PSNR measures the peak SNR in each band separately. Similarly, the SAM measures the spectral angle between the fusion product and the reference image for each pixel individually. Hence, PSNR and SAM are used to evaluate the fusion performance channel- and pixel-wise, respectively. Furthermore, both metrics are used as additional indicators for overall fusion accuracy by averaging the PSNR over all bands and the SAM over all pixels. Visual inspection of the fusion products and spatial error distributions is conducted by producing and comparing the following images: (1) RGB color-compositions, (2) pixel-wise

SAM, and (3) pixel-wise root mean square error (RMSE). Like the other aforementioned assessment measures, RMSE is defined in Section 2.2.4.

5.3.4 Hyperspectral-Multispectral Data Fusion Methods used for Comparison

For a fair and transparent evaluation of the fusion performance attainable with J-SparseFI-HM relative to the state of the art, the comparative assessment part of this chapter closely follows the recently published review article on hyperspectral-multispectral data fusion [Yokoya, Grohnfeldt, and Chanussot, 2017]. In that study, the following 10 partly established and partly recently proposed data fusion algorithms are selected as representative methods for different fusion categories: *Gram-Schmidt adaptive* (GSA) [Aiazzi et al., 2007] is selected as a representative component substitution (CS)-based method. Two multiresolution analysis (MRA)-based pan-sharpening methods, i.e., *smoothing filtered-based intensity modulation* (SFIM) [Liu, 2000] and *generalized Laplacian pyramid* (GLP) [Aiazzi et al., 2006], are adapted to the hyperspectral-multispectral data fusion problem via *hypersharpening* (HS) [Selva et al., 2015] – hence the name compositions SFIM-HS and GLP-HS. The *maximum a posteriori* estimation approach using a *stochastic mixing model* (MAP-SMM) [Eismann, 2004] and the *fast fusion based on Sylvester equation* (FUSE) algorithm [Wei et al., 2015c] are both based on Bayesian probability. For the latter algorithm, two versions are included in the comparative study: the "standard" version with Gaussian prior (FUSE) and a version with a sparse regularization prior (FUSE-S). The *coupled non-negative matrix factorization* (CNMF) [Yokoya et al., 2012] method, the *hyperspectral superresolution* (HySure) [Simões et al., 2015] method, *Akhtar's method* [Akhtar et al., 2014], and *Lanaras's method* [Lanaras et al., 2015] represent the fusion category of spectral unmixing-based approaches. Apart from the Bayesian-based methods, which both use PCA for subspace transformation, the unmixing-based approaches can be considered as subspace methods as well, because the dimension of the space spanned by the endmembers in these methods is generally smaller than the spectral dimension spanned by the unknown high-resolution hyperspectral data.

In order to simplify the experimental part of this chapter, only the best 7 out of 10 methods are included in the comparisons. Since GSA and Akthar's method showed the overall worst performance in the aforementioned review article, they are excluded from the experiments in this thesis. Moreover, the review article has confirmed the test results reported in [Wei et al., 2015c], which yielded that the sparse representation-based version of FUSE, i.e. FUSE-S, consistently produces better results in terms of product quality than the version based on a Gaussian prior. Hence, the latter version of FUSE is excluded from the experimental part of this thesis as well.

For more detailed information about the methodology of the algorithms under comparison, the interested reader is invited to refer to [Yokoya, Grohnfeldt, and Chanussot, 2017] or the respective original literature. MATLAB source codes of all algorithms under comparison are publicly available: The source codes of SFIM, GLP, and HySure are available on [OpenRemoteSensing, 2015, website]; The code of MAP-SMM is provided in [Eismann, 2004]; The software code of FUSE-S is available on [Wei, 2015, website]; The code of Lanaras' method can be accessed on [Lanaras, 2015, website] and the code of CNMF is available on [Yokoya, 2012, website].

In order to conduct a fair comparison, the fusion performance of each method was maximized via extensive parameter tuning and optimal algorithm setting for each data set individually as is described in detail in [Yokoya, Grohnfeldt, and Chanussot, 2017].

Fusion Method	Subspace	SRF	PSF	Reference
SFIM-HS	–	no	no	[Liu, 2000] + [Selva et al., 2015]
GLP-HS	–	no	yes	[Aiazzi et al., 2006] + [Selva et al., 2015]
MAP-SMM	PCA	no	no	[Eismann, 2004]
Lanaras'15	SISAL	yes	yes	[Lanaras et al., 2015]
FUSE-S	PCA	yes	yes	[Wei et al., 2015c]
CNMF	VCA	yes	yes	[Yokoya et al., 2012]
HySure	VCA/SVD	yes	yes	[Simões et al., 2015]

Table 7. Properties and literature references of the fusion methods used for comparison in this work. Subspace transformation methods used by the various fusion algorithms include *Principal Component Analysis* (PCA), *Simplex Identification via Split Augmented Lagrangian* (SISAL), *Vertex Component Analysis* (VCA) and *Singular Value Decomposition* (SVD). The second and third column indicate whether or not the algorithms requires the sensor's SRFs and PSFs, respectively, as input arguments.

Tab. 7 summarizes some properties which most of the seven fusion algorithms under comparison have in common. Those properties include information about which subspace projection method is used (if any) and whether or not the sensors' SRFs and PSFs are required as input. For the experiments discussed in this thesis, all methods requiring SRFs as input were run with the same SRFs, i.e. those estimated via Eq. (141). Similarly, those methods requiring PSFs as input (except for HySure) were run with the same Gaussian kernels as input, i.e. PSFs calculated via Eq. (13). Since HySure estimates the PSFs from the input data internally during the fusion process following the description in [Simões et al., 2015] and this does not appear to effect the fusion products negatively (see Section 5.5), the internally estimated PSFs are used for HySure instead of the Gaussian approximation.

5.3.5 Implementation and Computer Platform

The J-SparseFI-HM algorithm is exclusively implemented in C++. For linear algebraic operations, the `Eigen` template library [tuxfamily, 2017] (version 3.3.3. – stable release) is used. The software is optimized for processing on the SuperMUC petascale system, Phase 2 Haswell Nodes, which is operated by the Leibniz Supercomputing Centre (LRZ) [LRZ, 2017] in Garching (Munich), Germany. However, it can be easily installed on any common Unix-based system that supports the C++ libraries Eigen, MPI and GDAL (see descriptions below).

With more than 241,000 cores and a peak performance of more than 6.8 Petaflop/s, the SuperMUC was ranking 36th (Phase 1) and 37th (Phase 2) in the *Top 500* list of the most powerful commercially available computer systems in December 2016 [Top500, 2016]. One compute node consists of two processors, each constituting 14 cores. A single node can access 64 GByte of main memory, resulting in 2.3 GByte per core (typically, no more than 2.1 GByte is effectively useable). The Haswell Xeon E5-2697 v3 processors run at 2.6 GHz nominal frequency at best [SuperMUC, 2017].

Parallel processing is realized based on the `Intel MPI 2017 library`, which is Intel's latest implementation of the MPI 3.0 standard on SuperMUC. Internode MPI communication is kept low in order to maximize parallel efficiency. The J-SparseFI-HM application is compiled using the wrapper mpiCC.

Various parallelization strategies have been implemented and tested in the research towards this dissertation, including parallelization on the patch level (spatial), channel group level (spectral) and solver level using parallel FISTA (P-FISTA). Considering the most often significant difference between the spatial and spectral image dimensions, combined parallel processing of patches and channel groups has not shown advantages over the scenario where

the same number of cores were available for parallel patch processing; the latter scenario meaning that one core (or thread) may process multiple patches, but each patch is processed by a single core (or thread). Moreover, the parallelization of the sparse recovery algorithm FISTA, i.e. P-FISTA, has not shown advantages in realistic scenarios either. This result was to be expected, given that the number of sparse recovery problems is generally much larger than the dimension of the individual sparse recovery problems. Consequently, the most efficient parallelization approach was found to be the simple yet effective parallelization of patch reconstruction problems.

Regarding data I/O: in order to allow for direct processing of Earth observation data that may be provided in one of many common or uncommon data formats, a reading interface is implemented based upon the *Geospatial Data Abstraction Library* (GDAL) [GDAL, 2017] (version 2.2.0, stable release). In contrast to the reading interface, the output module does not use GDAL, but writes the potentially large fusion product file in parallel using MPI I/O. Parallel writing not only potentially accelerates the writing process but is useful when processing large data sets that may cause a memory bottle-neck when the full fusion product is written by – and therefore stored in the memory allocated to – a single core. The default output data format is implemented to be unsigned 16-bit integer, organized as band interleaved by pixel (BIP). In case higher precision is on demand, a simple flag parameter allows for switching the output format to 64-bit floating precision.

Depending on the dimension of the data to be processed, parameter setting and level of demand for fast processing, 4-8 nodes, i.e. 112-224 cores, typically finish processing between several minutes and a few hours. For the majority of experiments in this chapter, 5 nodes were used.

All other fusion algorithms were available as MATLAB programs, which were processed on a local computer featuring 32 GByte RAM and 8 Intel® Core^{TM} i7-6920HQ CPUs running at 2.90 GHz maximum frequency.

5.4 System and Parameter Analyses

This section presents a variety of numerical tests on the J-SparseFI-HM algorithm which complement the theoretical part described in Section 5.2. In particular, Section 5.4.1 discusses the estimation of sensor system quantities. Subsequently, the influence of the parameters of the local–non-local and global processing modules on the fusion performance is analysed in Section 5.4.2. Section 5.4.3 discusses processing times of both modules subject to parameter setting. Finally, the influence of the initial image $\boldsymbol{I}_{\tilde{Z}}$ on the fusion result and convergence rate is discussed in Section 5.4.4.

Due to its popularity in the remote sensing data fusion community [Debes et al., 2014] and the simulation of Sentinel-2 multispectral data, the first data set, i.e. CASI – *University of Houston*, was selected as underlying data set for the parameter analyses conducted in this section.

5.4.1 Estimation of Spectral Responses and Noise Variances

In Section 5.2.5, it is described how potentially unknown system characteristics including SFRs, PSFs and noise variances can be determined in a data-driven manner. While PSFs are approximated, in this work, by Gaussian functions with FHWM matching the GSD of the low-resolution sensor, the SRFs and noise variances are estimated from the data.

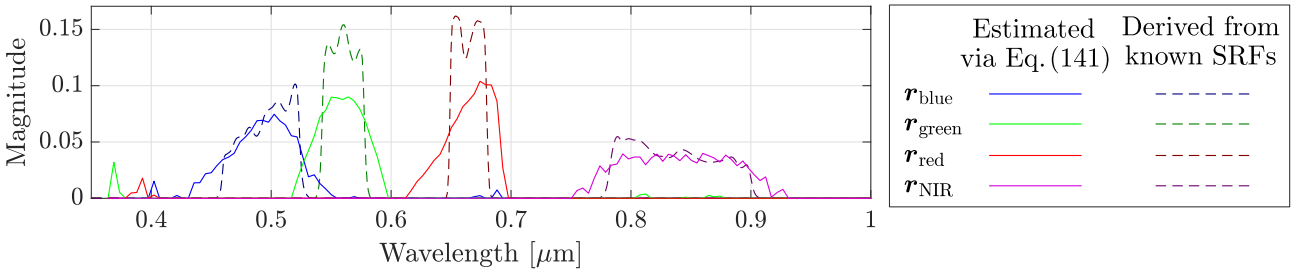


Fig. 58. Comparison of the results for the system matrix $\boldsymbol{R} = [\boldsymbol{r}_{\text{blue}}, \boldsymbol{r}_{\text{green}}, \boldsymbol{r}_{\text{red}}, \boldsymbol{r}_{\text{NIR}}]^T$ if estimated via Eq. (141) (solid lines) or derived from *a priori* known SRFs (dashed lines) for data set 1, i.e. CASI – *University of Houston*.

Spectral Response Functions The system matrix $\boldsymbol{R}$, which is introduced in the sensor observation model in Eqs. (99)–(100) and mainly constitutes information about the relative spectral responses of the multispectral imager, is estimated via the constrained non-negative linear regression problem stated in Eq. (141).

Fig. 58 shows a comparison of the results for the matrix $\boldsymbol{R} = [\boldsymbol{r}_{\text{blue}}, \boldsymbol{r}_{\text{green}}, \boldsymbol{r}_{\text{red}}, \boldsymbol{r}_{\text{NIR}}]^T$ if estimated via Eq. (141) (solid lines) or derived from *a priori* known SRFs (dashed lines). In the latter case, the known multispectral SRFs are filtered by the SRFs of the hyperspectral sensor, re-sampled to the corresponding hyperspectral SRF center wavelengths and finally normalized to fulfill the sum-to-one constraint (more detailed information is provided in Section 2.2.3). Since the latter matrix $\boldsymbol{R}$, which is derived from the original SRFs, is used for the simulation of the multispectral input image $\boldsymbol{I}_X$ before noise is added, it can be considered as the reference for the estimation result in the absence of additive noise. Fig. 58 reveals that even in the case of realistic SNR values of 35 dB, the estimation results do not differ significantly from the original SRFs. Over-estimated values in the 0.36-0.41 μm wavelength region can be expected considering the lower signal values in those bands which are a consequence of less solar irradiance reaching the Earth's surface (see Fig. 2 (c)). The spectral shift $\boldsymbol{t}$, which is computed along with $\boldsymbol{R}$ in Eq. (141), was estimated to be $t_{\text{blue}} = 81.22$, $t_{\text{green}} = 26.79$, $t_{\text{red}} = 41.77$, $t_{\text{NIR}} = 3.22$ for the underlying CASI data set. Expressed in percentages of the sample mean values of the corresponding multispectral channels (rows in $\boldsymbol{I}_X$), these numbers are equivalent to $t_{\text{rel,blue}} = 1.14\%$, $t_{\text{rel,green}} = 0.34\%$, $t_{\text{rel,red}} = 0.60\%$, $t_{\text{rel,NIR}} = 0.04\%$. Since no spectral shift was simulated explicitly, those marginal numbers are perfectly in line with what can be expected.

Noise Covariance Matrices $\boldsymbol{\Sigma_X}$, $\boldsymbol{\Sigma_Y}$ and $\boldsymbol{\Sigma_{\tilde{Z}}}$ In Section 5.2.5, it is explained how the choice of a suitable noise level estimation method depends on the spatial resolution and spectral information contained in the underlying data. Since the aim is to estimate the noise variances of both multi- and hyperspectral data, two noise level estimation methods, namely *Liu's method* [Liu et al., 2013] for high-resolution multispectral data and SSDC [Roger and Arnold, 1996] for hyperspectral data, are numerically tested and compared here. In order to evaluate the suitability of both estimation approaches for application to the underlying CASI data set, the information is used that the colored (i.e. band-dependent) noise was simulated to yield an SNR=35 dB in all channels. Reference signal variances σ_s^2 are available thanks to the availability of noise-free simulated multi- and hyperspectral input data, i.e. $\boldsymbol{I}_X$ and $\boldsymbol{I}_Y$ before the addition of noise. With σ_s^2 at hand, the estimation results for the noise variances σ_n^2 can be converted to SNR values via Eqs. (19)–(20) and directly assessed by comparing the resulting SNR values to the nominal constant value of SNR=35.

Fig. 59 shows the results of the experiment in which Liu's method and SSDC are both applied to (a) the hyperspectral input data $\boldsymbol{I}_Y$ and (b) the multispectral input data $\boldsymbol{I}_X$. The plots evidently confirm the theoretical argumentation from Section 5.2.5: Looking at Fig. 59 (a),

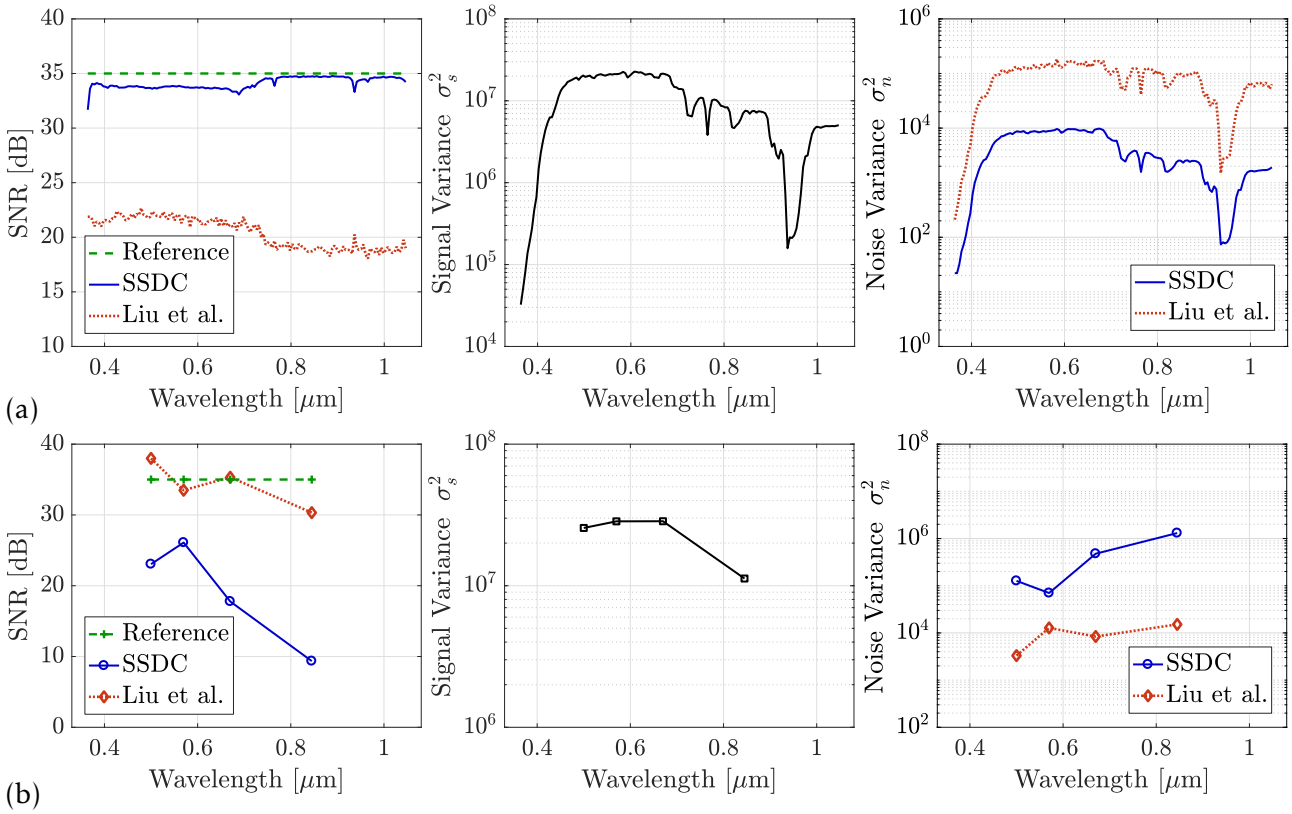


Fig. 59. Comparison of the noise level estimation results produced using SSDC [Roger and Arnold, 1996] and Liu's method [Liu et al., 2013] relative to the nominal constant SNR of 35 dB for (a) the hyperperspectral input data $\boldsymbol{I}_Y$ of the CASI data set and (b) the corresponding multispectral input data $\boldsymbol{I}_X$. From left to right, the figures show the SNR values, signal variances σ_s^2 and noise variances σ_n^2 over the wavelength. The data points on the horizontal axes correspond to the centers of the channels' SRFs.

the SSDC method, which is designed for the noise level estimation of *hyperspectral* data, yields SNR values that are remarkably close to the nominal 35 dB especially in the NIR channels. In contrast, Liu's method, which is a single-image noise level estimation approach and, therefore, does not take into account spectral information, severely over-estimates the noise variance in the data. This result is to be expected considering that Liu's method requires finding square homogeneous areas in the image. Like in most *low-resolution* images acquired over urban (heterogeneous) ground, homogeneous areas are rare to find in this scene (see Fig. 55). Moreover, with a GSD ratio of 5, the underlying hyperspectral image $\boldsymbol{I}_Y$ contains 25 times fewer pixels than the high-resolution multispectral input image $\boldsymbol{I}_X$. Hence, SSDC is clearly favorable for noise level estimation for hyperspectral data and therefore used to determine $\boldsymbol{\Sigma}_{\boldsymbol{Y}}$ and $\boldsymbol{\Sigma}_{\tilde{\boldsymbol{Z}}}$. On the other hand, when looking at the estimation results for the high-resolution multispectral input data in Fig. 59 (b), one sees the exact opposite performances of the two algorithms under comparison. The limited spectral information in $\boldsymbol{I}_X$ makes SSDC over-estimate the noise level in each channel. However, the higher spatial resolution makes Liu's method estimate the nominal 35 dB SNR more accurately and is, therefore, the better choice for the estimation of $\boldsymbol{\Sigma}_X$.

5.4.2 Parameters Analysis

The full list of parameters with corresponding values used for the processing of all eight data sets in this chapter is presented in Tab. 8. That table also indicates which parameters belong to the local–non-local processing module and which belong to the global processing

	Local–non-local processing module							Global processing module	
	N_{L}^{P}	λ	N_a	θ	$\hat{N}_c$	μ_X	μ_Y	μ_Z'	$N_{b,\mathrm{sub}}^{Y}$
CASI – *University of Houston*	4^2	10^1	1000	0.9	40	$10^{-1/2}$	$10^{-1/2}$	$10^{-1/4}$	13
Hyperspec – *Chikusei*	4^2	$10^{1/2}$	900	0.97	15	$10^{1/2}$	10^0	10^0	17
ROSIS-3 – *University of Pavia*	3^2	10^0	900	0.96	25	$10^{1/2}$	$10^{-1/2}$	10^0	10
HYDICE – *Washington D.C. Mall*	4^2	10^0	800	0.95	25	$10^{-3/2}$	10^0	10^0	18
AVIRIS – *Indian Pines*	4^2	10^0	1000	0.85	30	$10^{-3/2}$	10^{-1}	$10^{3/4}$	14
AVIRIS – *Cuprite*	4^2	10^{-1}	900	0.9	25	$10^{-5/2}$	10^{-2}	10^0	15
AVIRIS – *Moffett Field*	4^2	$10^{1/2}$	650	0.9	30	10^{-1}	10^0	10^0	18
HyMap – *Rodalquilar*	4^2	$10^{1/2}$	300	0.99	15	10^0	$10^{-1/2}$	$10^{1/8}$	14

Table 8. List of parameters and corresponding values used for the fusion of the eight data sets in the experimental part of this chapter. From left to right, the nine parameter symbols can be written out as follows: N_{L}^{P}: patch size; λ: sparsity regularization parameter; N_a: dictionary size; θ: correlation threshold; $\hat{N}_c$: lower bound for large spectral groups; μ_X/μ_Y: relative contribution of the multispectral/hyperspectral input data $\boldsymbol{I}_Y/\boldsymbol{I}_X$ to the joint sparse coefficient estimation; μ_Z': relative contribution of the intermediate fusion product $\boldsymbol{I}_{\widetilde{Z}}$ to the estimation result of the global image coefficients $\boldsymbol{J}_Z$; $N_{b,\mathrm{sub}}^{Y}$: hyperspectral subspace dimension.

module. In the following, the impact of each adjustment parameter on the fusion performance is exemplarily tested and briefly discussed.

As is further discussed in Section 5.4.4, from a computational point of view it is advantageous to use an efficient fusion algorithm to produce an initial image $\boldsymbol{I}_{\widetilde{Z}}$ instead of, e.g., a bilinearly interpolated and sub-sampled version of the hyperspectral input image $\boldsymbol{I}_Y$. The parameter analyses conducted in this section are based on the fusion result of the CASI data set produced by CNMF. Following the flowchart of J-SparseFI-HM depicted in Fig. 51, this means that, in the first iteration, the local–non-local processing module uses the CNMF product as input for $\boldsymbol{I}_{\widetilde{Z}}$ whereas the global processing module uses the output of the local–non-local processing module. The following parameter tests indicate the quantitative performance in PSNR, ERGAS and SAM after one iteration. In the tests of the parameters corresponding to the first (local–non-local) module, the output of that module – i.e. the intermediate fusion product – is assessed. In the tests of the remaining two parameters μ_Z' and $N_{b,\mathrm{sub}}^{Y}$, the output of the second (global) processing module is assessed. While the variation of each parameter is discussed individually, all remaining parameters are fixed to the values shown in Tab. 8, line 3, CASI – *University of Houston*.

Patch Size N_{L}^{P} Fig. 60 shows the impact of both the patch size N_{L}^{P} and patch overlap N_{L}^{PO} on the overall quality of the fusion product assessed via PSNR, ERGAS and SAM. Since patches are set square and specified in numbers of low-resolution pixels, a patch size of, e.g., $N_{\mathrm{L}}^{P} = 4^2$ implicates that all patches are of size 4×4 low-resolution or 20×20 high-resolution pixels since the GSD ratio is equal to 5. For each of the patch sizes $N_{\mathrm{L}}^{P} = 2^2, 3^2, \ldots, 7^2$ tested, the full range of possible numbers for patch overlap is testes as well. That is, $N_{\mathrm{L}}^{PO} = 0, 1, \ldots, \sqrt{N_{\mathrm{L}}^{P}} - 1$. As a first result, Fig. 60 evidently confirms the observation that has been made in the previous chapter on J-SparseFI already: A maximum overlap of $N_{\mathrm{L}}^{PO} = \sqrt{N_{\mathrm{L}}^{P}} - 1$ reliably yields the best reconstruction results independently of the size of the patches. Hence, maximum patch overlap is recommended to be used, which is also followed in the remainder of this chapter. As for the patch size, there is no number which yields best results in terms of all assessment metrics. The PSNR and SAM plots suggest using a patch size of $N_{\mathrm{L}}^{P} = 4^2$ whereas the best ERGAS value is achieved with the larger patch size of $N_{\mathrm{L}}^{P} = 6^2$. Trade-offs are unavoidable in parameter setting as will be observable

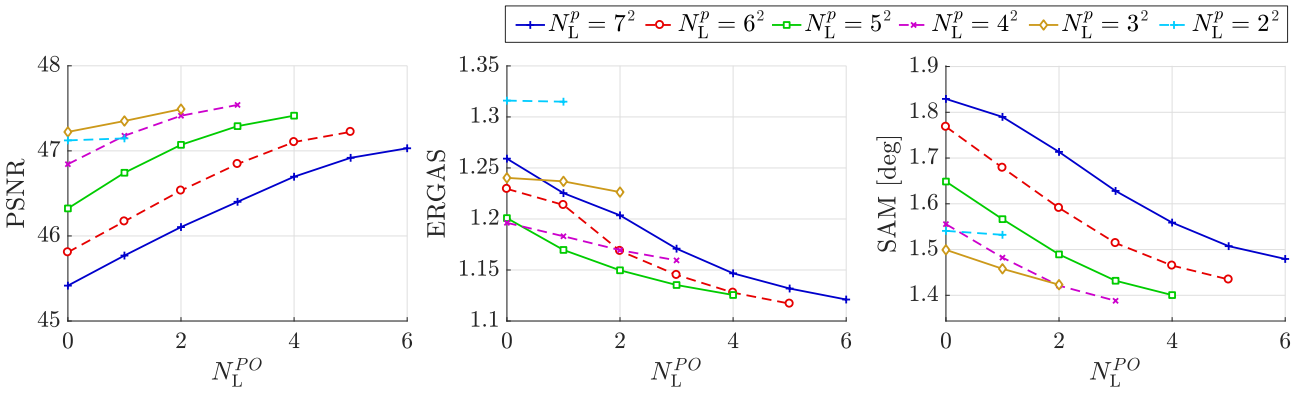


Fig. 60. Impact of both the patch size N_L^P and patch overlap N_L^{PO} on the quality of the J-SparseFI-HM fusion product assessed via PSNR (left), ERGAS (center) and SAM (right).

in many of the following experiments. Taking into account considerable experience gained from tests on other data sets, the GSD ratio plays an important role and should be consulted for finding the patch size best suitable for the data set at hand. This is because, the size of the *high-resolution* patches increases *quadratically* in both $\sqrt{N_L^P}$ and the GSD ratio. Tab. 8 shows that the only data set for which a patch size of 3^2 is selected instead of 4^2 is ROSIS-3 – *University of Pavia*, which has the largest GSD ratio of 8.

Sparsity Regularization Parameter λ and Dictionary size N_a The (joint) sparsity-regularizing parameter λ and the dictionary size N_a are the only two parameters that affect only the non-local component of the local–non-local processing module. They are used in steps ④ b) and c) after the local coefficients $\boldsymbol{\alpha}_i^p$ corresponding to the local patch (first atom in the coupled dictionary pair) are estimated in step ④ a) via Eq. (125).

Fig. 61 (a) and (b) illustrate the impact of λ and N_a, respectively, on the fusion performance. Moreover, Fig. 61 (b) shows the different performances achieved using the four dictionary atom selection approaches described in Section 5.2.3 (step ③ b) of J-SparseFI-HM).

In Fig. 61 (a), the peak in the PSNR plot and the troughs in both the ERGAS and SAM plots around $\lambda = 10$ clearly indicate the performance improvement brought by exploiting (joint) sparsity. As λ increases, the non-local coefficients $\hat{\boldsymbol{A}}_{\text{res},i}^p$ become sparser and sparser. For $\lambda > 10^3$, the performance remains unchanged which indicates that the coefficients have reached zero. This limit is identical with the results attained by setting $N_a = 1$, i.e. using the trivial dictionary that contains only the local patch. As $\lambda \to 0$, $\hat{\boldsymbol{A}}_{\text{res},i}^p$ becomes denser. In terms of ERGAS, the latter case even worsens the product created using only the local patch reconstruction, i.e. the coefficients $\boldsymbol{\alpha}_i^p$. Optimim values for λ may slightly vary from one data set to another, even though not significantly thanks to the normalization of dictionary atoms. Tab. 8 presents the optimum values found for the various data sets investigated in this thesis. Those values all lie between 10^{-1} and 10^1, where the relatively small value of 10^{-1} is only used for the Cuprite data set which is comparatively noisy. For all other data sets, values for λ between 1 and 10 were found to be optimal, which is the recommended setting for unknown data sets that do not severely suffer from low SNRs.

In Fig. 61 (b), the plots corresponding to all four selection methods meet at dictionary size of $N_a = 6405$, which is the total number of available patches in the underlying data set at the default patch size of $N_L^P = 4^2$ and maximum patch overlap. Apart from this, the graphs allow two observations:

(1) The nearest neighbor selection method, which is indicated by μ_{NN} and used for the

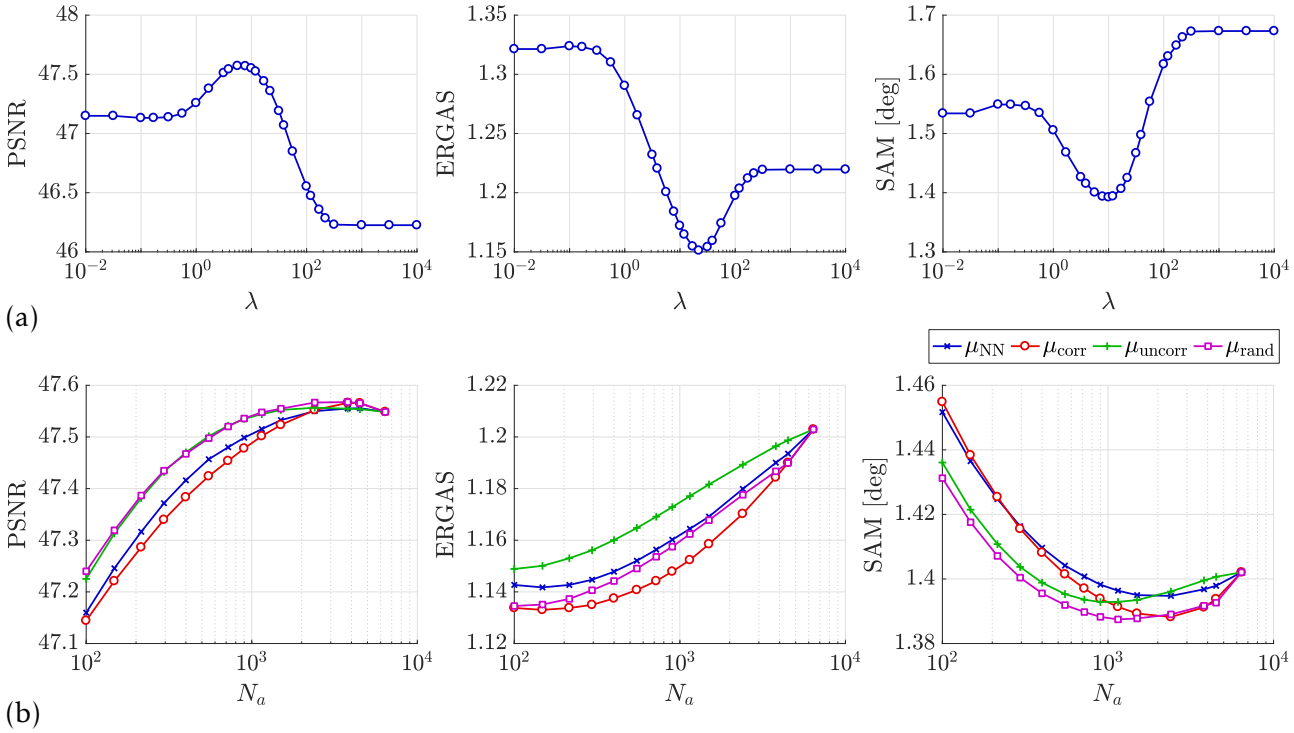


Fig. 61. Impact of (a) the (joint) sparsity regularizing parameter λ and (b) the dictionary size N_a on the fusion performance attained via the local–non-local processing module of J-SparseFI-HM.

J-SparseFI pan-sharpening algorithm in Chapter 4, performs average in terms of PSNR and ERGAS and worst in terms of SAM. The selection approach performing best in both PSNR and SAM and second best in ERGAS is the simple yet effective random atom selection approach. Similar observations have been made in various tests on different data sets, which is why this approach is generally recommended for usage and will be used in the remainder of this chapter.

(2) The dictionary size N_a affects the fusion performance attained via all four atom selection approaches in a similar manner. With a view to the plots corresponding to the random selection method, indicated by μ_{rand}, one can clearly see the afore-mentioned trade-off in performance between ERGAS and the other two assessment measures. For all sample data sets used in this work, N_a is set to moderate numbers between 650 and 1000. Considering the relatively low sensitivity of the performance to chances in N_a within that number range, any number in between, such as $N_a = 800$, is save for recommendation.

Correlation Threshold θ and Lower Bound for Large Spectral Groups $\hat{N}_c$ Fig. 62 (a) shows the performance improvement gained through the proposed Correlation-based HyperSpectral Grouping (CorHySpeG) concept and the subsequent joint processing of mutually correlated channels. In particular, the results attained with $\theta = 1$ (right-most data point in each of the three sub-figures) represents the scenario where each channel is processed separately. In this case, each band has a corresponding individually sparse coefficient vector. The smaller θ the lower is the minimum required mutual correlation between channels in the same group. In other words, the smaller θ the larger the groups. The curves colored in green, red and blue illustrate the scenarios, in which the lower bound for large groups is set to $\hat{N}_c = \infty$, 40, and 5, respectively.

The first case, i.e. $\hat{N}_c = \infty$, is interesting as it represents the scenario where no post-consideration of large groups and, therefore, no post-adaptation of θ is conducted. With

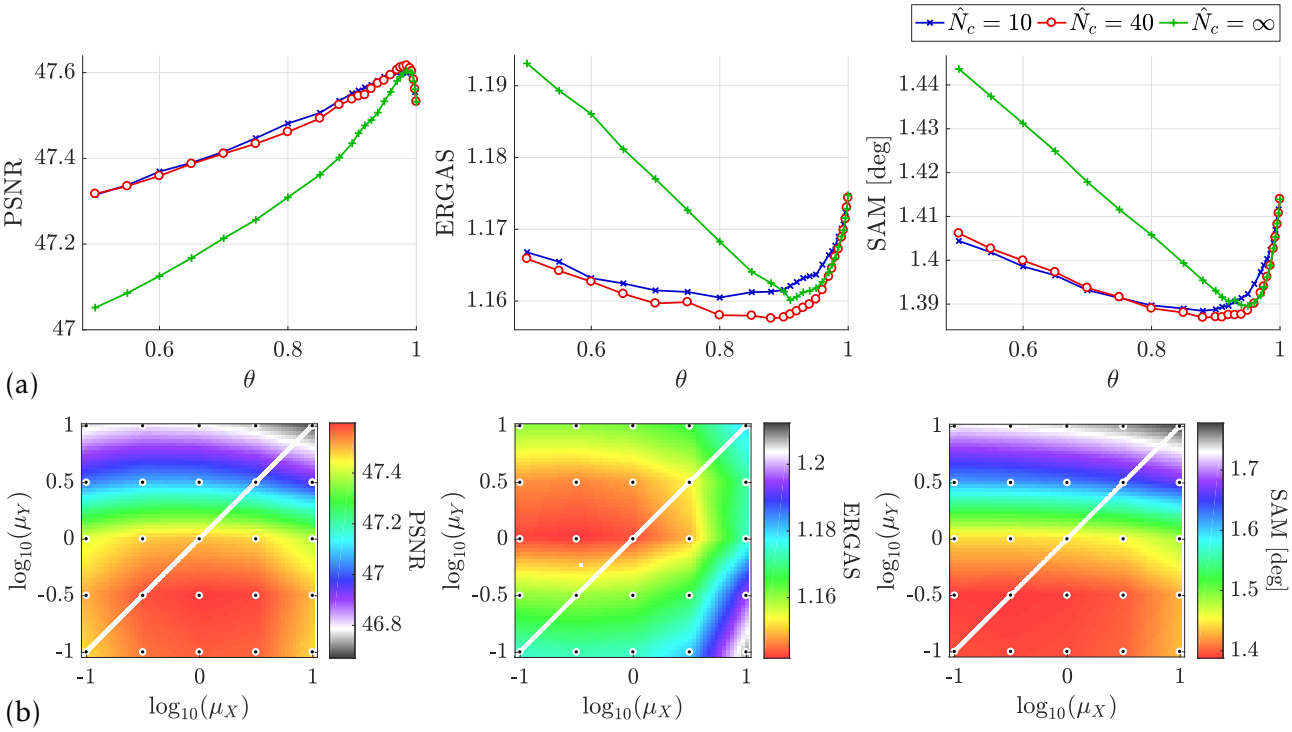


Fig. 62. (a) Impact of the correlation threshold θ and the lower bound for large channel groups on the performance of the local–non-local processing module of J-SparseFI-HM. The figure indicates the fusion quality improvement achieved through the proposed Correlation-based HyperSpectral Grouping (CorHySpeG). (b) Influence of the weights μ_X and μ_Y on the quality of the output of the local–non-local processing module.

a view to the flowchart of CorHySpeG shown in Fig. 52, this scenario occurs if steps 5 and 6 are skipped. In this case, the performance is much more sensitive to the original choice of θ and generally lower compared to the case in which θ is automatically post-adjusted for each large channel group.

If $\hat{N}_c$ is set too small (here, represented by the case of $\hat{N}_c = 10$), the constraint on the size of the spectral channel groups may cause loss of performance, too, as can be seen in Fig. 62 (a). If no reference data is available, $\hat{N}_c$ is recommended to be set relative to the spectral sampling distance and the spectral signatures of the materials of interest in the scene. Green vegetation spectra, for instance, feature comparatively broad plateaus within which channels can be expected to be highly correlated. If, in addition to that, the sampling distance of the hyperspectral imager is fine, N_c should be set larger. In most of the experiments included in this work, moderate numbers for $\hat{N}_c$ of either 25 or 30 yield stable performance improvements. The low spectral sampling distance of HyMap and the comparatively fine spectral sampling distance of CASI explain the deviating numbers for $\hat{N}_c$ in Tab. 8.

A suitable value for the correlation threshold θ can be found by analysing the global spectral correlation matrix of either the data to be processed or a representative data set acquired by the same hyperspectral sensor. Correlation matrices such as those shown in Figs. 22 and 52, usually give clearly indication of the level of mutual channel correlation inherent in the hyperspectral data.

Relative Contributions μ_X and μ_Y of the Input Data $\boldsymbol{I}_X$ and $\boldsymbol{I}_Y$ Fig. 62 (b) shows the influence of the weights μ_X and μ_Y on the quality of the output of the local–non-local processing module. The gray lines along the diagonals of the three sub-figures indicate the linear parameter subspaces corresponding to the solutions achievable by setting $\mu_X = \mu_Y$. This case is interesting as it indicates that the reduction of those two parameters to a single one may

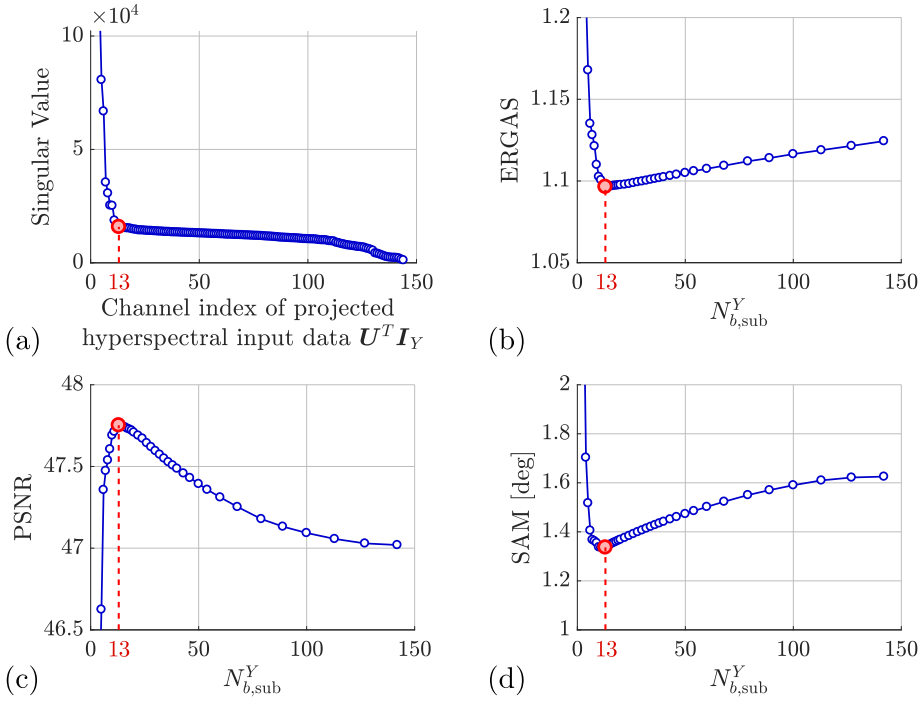


Fig. 63. (a) Sorted singular values of the input low-resolution hyperspectral data $\boldsymbol{I}_Y$. (b)–(d) Influence of the spectral subspace dimension $N^Y_{b,\text{sub}}$ on the quality of the fusion product measured via (b) ERGAS, (c) PSNR and (d) SAM. In all four sub-figures, the intrinsic spectral dimension of $N^Y_{b,\text{sub}} = 13$ is indicated.

not cause a significant loss of performance in terms of quality. Note that, if μ_X and μ_Y are replaced by a single parameter μ_Z, which weighs the contribution of the initial imge $\boldsymbol{I}_{\tilde{Z}}$ in Eqs. (106), (125), (128) and (130), the problem would be slightly modified in the following sense: In order for the two problems to be equivalent, the first term in Eq. (128) would need to multiplied by μ_Z, which would not serve the purpose of μ_Z, i.e. weighing the contribution of $\boldsymbol{I}_{\tilde{Z}}$. Hence, the new solution subspace would be slightly different to what is displayed in Fig. 62 (b). Nonetheless, the reasonable trade-off between PSNR, ERGAS and SAM quality measures in the range $\mu_X = \mu_Y \in [10^{-1/2}, 10^0]$ suggests that, first, the normalization factors before the corresponding Frobenious terms are effective, and, second, parameter reduction is worth to consider in the future. The latter topic is discussed as a potential future work in Chapter 6. In the case of the CASI data set underlying this analysis, the parameters μ_X and μ_Y are both set to $10^{-1/2}$.

Since μ_X and μ_Y weigh the impact of the two input images $\boldsymbol{I}_X$ and $\boldsymbol{I}_Y$ relative to the initial image $\boldsymbol{I}_{\tilde{Z}}$, they should be set subject to the estimated quality of $\boldsymbol{I}_{\tilde{Z}}$. As is discussed further below, if $\boldsymbol{I}_{\tilde{Z}}$ is set to a bilinearly interpolated and sub-sampled version of the hyperspectral input data instead of a presumably more accurate estimate, the optimum range of the parameter space spanned by μ_X and μ_Y is slightly shifted towards greater values.

Spectral Subspace Dimension $N^Y_{b,\text{sub}}$ Figs. 63 (b) through (d) show the influence of the spectral subspace dimension $N^Y_{b,\text{sub}}$ on the quality of the fusion product measured via (b) ERGAS, (c) PSNR and (d) SAM. Fig. 63 (a) depicts a plot of the sorted singular values of the input low-resolution hyperspectral data $\boldsymbol{I}_Y$. In all four sub-figures, the intrinsic spectral dimension of $N^Y_{b,\text{sub}} = 13$ is indicated. That number can be derived directly and reliably from original hyperspectral data by analysing, for instance, the singular values of $\boldsymbol{I}_Y$. Other means of identifying $N^Y_{b,\text{sub}}$ are described in, e.g., [Bioucas-Dias and Nascimento, 2008; Torti et al., 2014; Rasti et al., 2015a; Benner et al., 2015]. Clearly visible is the coincidence of the

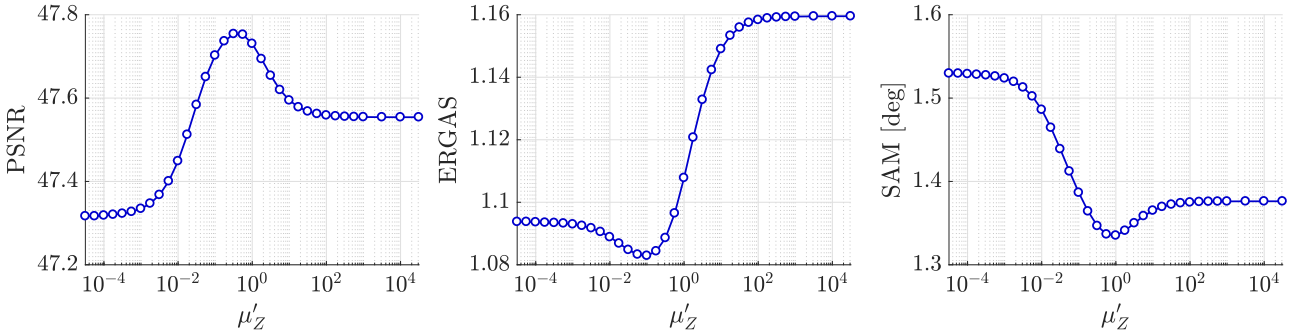


Fig. 64. Illustration of the fusion quality improvement achievable through the global processing module depending on the parameter $\boldsymbol{I}_{\tilde{Z}}$. That parameter weighs the contribution of the output of the local–non-local processing module, i.e. the intermediate fusion result $\boldsymbol{I}_{\tilde{Z}}$, relative to the impact of the original input data $\boldsymbol{I}_X$ and $\boldsymbol{I}_Y$ in Eq. (138).

intrinsic spectral dimension with the data point where peak quality is attained in terms of all three assessment metrics. In addition to that coincidence, the figure reveals the following information about the impact of conducting optimization on an SVD-based spectral subspace.

(1) If the subspace dimension is set too small, which, for the underlying CASI data, is the case if $N^Y_{b,\text{sub}} \leq 8$, the amount and importance of the signal information that is lost through dimensionality reduction becomes more and more significant, which causes a rapid drop of the fusion product's quality.
(2) As the subspace dimension $N^Y_{b,\text{sub}}$ exceeds the true intrinsic spectral dimensionality, the signal level in the SVD-projected data $\boldsymbol{U}^T\boldsymbol{I}_Y$ rapidly decreases while the noise level increases. Hence, the denoising capacity of SVD is best effective if $N^Y_{b,\text{sub}}$ is set to number not much larger than the true intrinsic spectral dimensionality.

Regularization Parameter μ'_Z The parameter μ'_Z regulates the contribution of the output of the local–non-local processing module, i.e. the intermediate fusion result $\boldsymbol{I}_{\tilde{Z}}$, relative to the impact of the original input data $\boldsymbol{I}_X$ and $\boldsymbol{I}_Y$ in Eq. (138). Thanks to the use of estimated noise variances as channel-specific weights, performance curves over μ'_Z – such as those shown in Fig. 64 – are nearly identical for all data sets. Looking at Tab. 8, one sees that optimum values are found at $\mu'_Z = 1$ for almost every data set. If the output of the local–non-local processing module is expected to be of lower quality, μ'_Z should be set to a smaller value, accordingly.

5.4.3 Computational Time

Fig. 65 shows the dependency of the processing time on (a) the sparsity regularization parameter λ, (b) the dictionary size N_a, (c) the regularization parameter μ'_Z, and (d) the spectral subspace dimension $N^Y_{b,\text{sub}}$. Sub-figures (a) and (b) display the total times taken by the local–non-local processing module using 5 nodes (140 cores) whereas sub-figures (c) and (d) refer to the processing times taken by the global processing module using 1 node (28 cores). The times displayed in sub-figures (a), (b), (c) and (d) were measured during the tests whose quality assessment results are shown in Fig. 61 (a), Fig. 61 (b), Fig. 64 and Fig. 63 ((b)-(d)), respectively.

Fig. 65 (a) clearly reveals the massive impact of the sparse recovery component of the J-SparseFI-HM algorithm formulated in Eq. (128) on the overall processing time. For $\lambda \gtrapprox 10^2$, the (joint) sparse recovery solver (J-)FISTA takes only 1 iteration to find the triv-

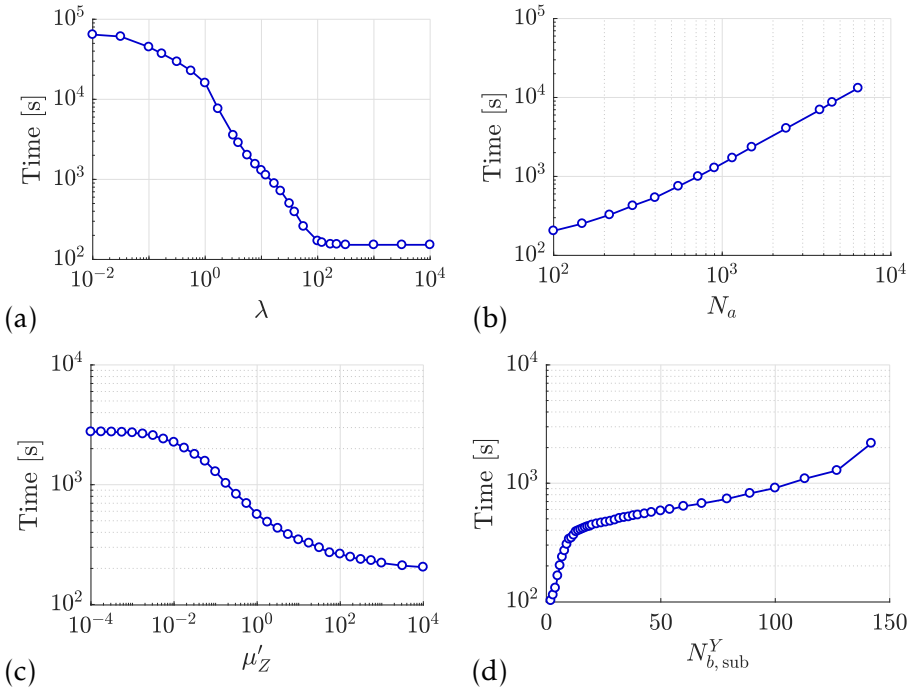


Fig. 65. Dependency of the processing time on (a) the sparsity regularization parameter λ, (b) the dictionary size N_a, (c) the regularization parameter μ'_Z, and (d) the spectral subspace dimension. Sub-figures (a) and (b) display the total times taken by the local–non-local processing module using 5 nodes (140 cores) whereas sub-figures (c) and (d) refer to the processing times taken by the global processing module using 1 node (28 cores).

ial solution $\hat{\boldsymbol{A}}^{\text{aux},p}_{\text{res},i} = \boldsymbol{0}$. This means that the processing time for $\lambda \to \infty$ of 152 seconds ($\approx$ 2.5 minutes) corresponds to the processing time of the local–non-local processing module without Eqs. (128) and (130). As λ is set to 10 for the tests on the CASI data set (see Tab. 8), the first module takes ca. 10^3 seconds ($\approx$ 17 minutes), for one iteration. Hence, the non-local sparse recovery part of the first processing module is responsible for the overhead of 14.5 minutes in this example.

Fig. 65 (b) shows the time increments the local–non-local processing module takes for processing with increasing dictionary sizes. Along the lines of the argumentations provided for the time dependency on λ, the nearly linear increase of the processing time with growing dictionary sizes is, primarily, a consequence of the increasing computational complexity of FISTA. In particular, the most demanding part in the joint sparsity-adapted version of Alg. 2 (FISTA) are the multiplications $\boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{X}$ and $\boldsymbol{\Phi}^T\boldsymbol{Y}$ which need to be evaluated in every iteration of FISTA. Since $\boldsymbol{\Phi} = [\boldsymbol{D}^{p}_{\text{res,H},i}{}^T, \boldsymbol{D}^{p}_{\text{res,L},i}{}^T]^T$ in Eq. (128), the increasing number of operations that come with an increase of the number N_a of dictionary atoms, neatly explains the development of the curve shown in Fig. 65 (b). As $N_a \to 1$, the processing time can be expected to approach a value very similar to the 152 seconds observed above for $\lambda \to \infty$, except for a marginal additional time the dictionary generation step takes if $N_a > 1$.

Interesting to observe is the significant correlation between the results on the time dependencies on λ and N_a which are obtained, individually, for J-SparseFI (see Fig. 44 (a) and (b)) and J-SparseFI-HM (see Fig. 65 (a) and (b)).

Fig. 65 (c) displays a representative plot of the processing time against the regularization parameter μ'_Z. The curve reveals that the processing time of the global processing module monotonically decreases as the influence of the first fidelity term in Eq. (138) increases. This development is conform to the intuition that finding coefficients $\boldsymbol{J}_Z$ that make the SVD basis stored in $\boldsymbol{W}$ approximate $\boldsymbol{I}_{\tilde{Z}}$ best – i.e. minimizing the first term in Eq. (138) – is numerically

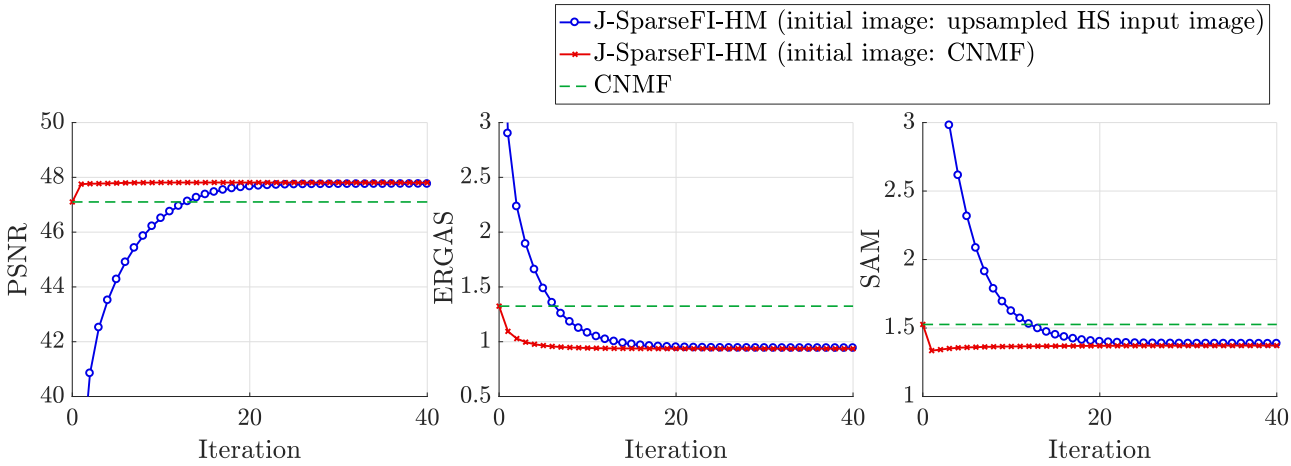


Fig. 66. Impact of the initial image on the convergence rate of J-SparseFI-HM. Shown is the propagation of the reconstruction error against the iteration number of J-SparseFI-HM measured via PSNR (left), ERGAS (center) and SAM (right). The relatively accurate initial image produced via CNMF yields convergence after a few iterations (red line) whereas a very coarse initial image obtained via bilinear interpolation and subsampling of the input hyperspectral image $\boldsymbol{I}_Y$ entails a slower convergence rate (blue line).

less challenging than finding coefficients $\boldsymbol{J}_Z$ that simultaneously satisfy the second and third fidelity terms in Eq. (138).

Finally, the steady increase of the processing time with growing $N^Y_{b,\text{sub}}$ is a simple consequence of the increasing number of coefficients in $\boldsymbol{J}_Z$ which the cost functional in Eq. (138) is optimized for. The bending of the curve around the aforementioned intrinsic spectral dimensionality of $N^Y_{b,\text{sub}} = 13$ indicates the relatively abrupt decrease of signal information, which is mainly contained in the first few basis vectors in $\boldsymbol{W}$.

5.4.4 Initial Image and Convergence

Fig. 66 shows the impact of the initial image on the convergence rate of J-SparseFI-HM. Shown is the propagation of the reconstruction error against the iteration number of J-SparseFI-HM measured via PSNR (left), ERGAS (center) and SAM (right). The relatively accurate initial image produced via CNMF yields convergence after a few iterations (red line) whereas a very coarse initial image obtained via bilinear interpolation and subsampling of the input hyperspectral image $\boldsymbol{I}_Y$ entails a slower convergence rate (blue line). The latter observation is a simple consequence of the algorithm being designed to create, in each iteration, an estimate $\boldsymbol{I}_{\hat{Z}}$ that satisfies the sensor observation model, i.e. Eqs. (99)–(100), while not deviating too much from the product of the previous iteration.

Considering the computational efficiency of CNMF (see e.g. [Loncan et al., 2015]) and its robustness in terms of quality (see Section 5.5), it is recommended to be consulted as initializer for $\boldsymbol{I}_{\tilde{Z}}$. Currently undergoing investigation is the automation of J-SparseFI-HM towards independently and efficiently estimating an accurate initial image in the first iteration. The latter topic is elaborated on in the discussion on future work in Chapter 6.

5.5 Fusion Quality Assessment and Comparison to the State of the Art

The quality of the data fusion products is evaluated using the complementary assessment criteria described in Section 5.3.3. In particular, Section 5.5.1 provides, for each of the eight data sets introduced in Section 5.3.1, a visual analysis of the J-SparseFI-HM high-resolution hyperspectral data fusion product relative to the input and reference data. Moreover, it presents illustrations and discussions of the spatial and spectral error distributions in the data produced via each method under comparison including J-SparseFI-HM. Section 5.5.2 presents a comparison of the overall fusion accuracy measured, for each method and data set, via various quantitative assessment metrics. As a practical demonstration of the versatility and usefulness of the J-SparseFI-HM data fusion products, Section 5.5.3 presents a classification experiment which evidences the algorithm's capacity for classification accuracy improvement.

5.5.1 Visual, Pixel-wise and Band-wise Evaluation

Figs. 67, 69, 71, 73, 75, 76, 77, 78, depict, for each of the eight data sets, color-composite images of the input multispectral data, the input hyperspectral data, the J-SparseFI-HM fusion product and the reference high-resolution hyperspectral data. The multispectral images are all shown as true-color composites whereas the hyperspectral images are visualised using false-color RGB channel composites corresponding to wavelengths that are *not covered* by the respective multispectral sensor. This way, the resolution enhancement capacity of J-SparseFI-HM is evidently visible, especially if a multispectral sensor featuring only VNIR bands is used for sharpening hyperspectral channels in the SWIR portion of the spectrum. The wavelengths selected to represent red, green and blue differ from one data set to another and are specified in the captions of the corresponding figures.

As differences between the quality of fusion products obtained via different fusion algorithms are visually much more pronounced in error maps – showing the product's spatial error distribution relative to reference data – than in color-composite images, the comparative evaluation of all fusion methods is based on pixel-wise computed spectral angles and root mean square errors. In addition to error maps, the wavelength-dependent performance is analysed using plots of the PSNR – as computed for each band – against the wavelength. The latter means of assessment is important, because the error distribution across the hyperspectral channels has a major impact on many applications that are based on spectral indices and band ratios.

In order to facilitate visual analyses, the color composite images and error maps shown in this section are spatially cropped to be of square shape with maximum spatial extends for each data set. The band-wise and overall quality assessment results are based on the original-size data whose spatial extends are specified in Tab. 6 and depicted in Fig. 55.

To avoid unnecessary repetition, only four out of eight data sets, namely the four most distinct ones in terms of spectral and spatial characteristics, are used for comparative error analyses. Those data sets include CASI (*University of Houston*), ROSIS-3 (*Pavia*), AVIRIS (*Cuprite*) and HyMap (*Radalquilar*). The corresponding error maps and PSNR plots are shown in Figs. 68, 70, 72 and 74, respectively.

In the following paragraphs, the results obtained for each of those four data sets are de-

scribed and discussed individually. After that, the conclusions drawn from those discussions are corroborated by additional color-composite images of the fusion products obtained for the remaining four data sets. Overall quantitative quality assessment is conducted, equally, for all eight data sets in Section 5.5.2.

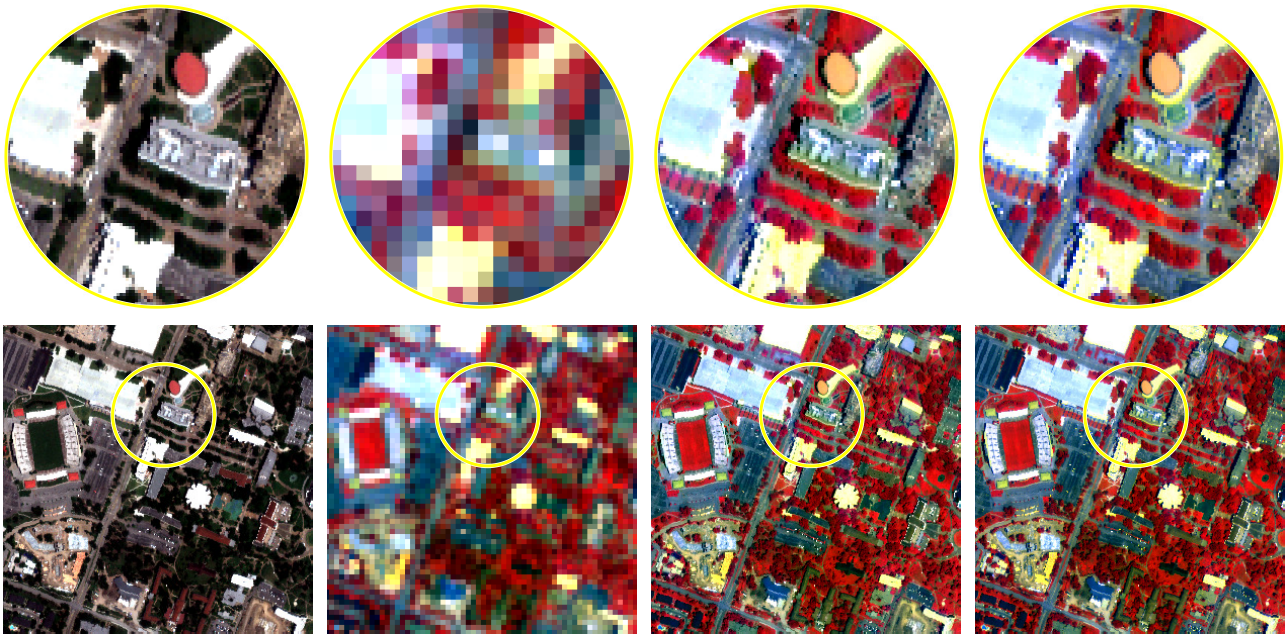
CASI – *University of Houston* Fig. 67 illustrates the *overwhelming level of reconstruction accuracy* J-SparseFI-HM is capable of achieving. The very high quality of the fusion product is evident when inspecting, first, the input images in comparison to the fusion product, then the magnified views above each color-composite image and, finally, the visually barely noticeable difference between the fusion product and the reference image. Thanks to the combined exploitation of local, non-local and global information in the fusion process, the preservation of spatial details from the multispectral input data appears nearly perfect. The only noticeable differences between the fusion result and reference data are little deviations in the color (e.g. different tones of yellow) of spatially isolated objects or along edges.

A more substantiated error analysis can be made based on Fig. 68. The spatial error distributions shown in Figs. 68 (a) and (b) indicate the strong competitiveness of J-SparseFI-HM to the state-of-the art in hyperspectral-multispectral data fusion. Recall that the seven algorithms used for comparison here were found to be the *best-performing* fusion methods out of ten partially recently published and partially well-established algorithms which are thoroughly tested and compared in the comprehensive review article on that subject recently published in [Yokoya et al., 2017].

In Fig. 68 (a), the low level of spectral distortion of the J-SparseFI-HM product, relative to the results obtained with the other seven methods, is distinctive mostly in the vegetated parts of the scene (see Fig. 67). Additional information on the spatial error distribution can be found in Fig. 68 (b): The only method which shows, in some regions, lower RMSE values than J-SparseFI-HM is HySure, i.e., the method ranked best in [Yokoya et al., 2017]. Similar observations can by made when analysing the results obtained with other data sets shown below. The explanation for this is fairly simple: HySure uses *total variation* (TV) to regularize an optimization problem similar to the one stated in Eq. (138). The minimization of gradients in the spatial domain makes HySure feature excellent denoising capacity. This property shows its strength particularly in *homogeneous areas*. As can be clearly seen when comparing Fig. 67 and Fig. 68 (b), the areas in which the HySure product shows noticeably lower RMSE values than the J-SpareseFI-HM product correspond, indeed, to homogeneous areas such as parking lots and planar building roofs. It is important to note here, that the denoising property attained through TV regularization comes at the expense of performance in detecting isolated materials, which is crucial in many hyperspectral remote sensing applications.

The data fusion quality measured for each band individually, is shown in Fig. 68 (c). This plot provides information about the performance of the methods under comparison, which is complementary to the observations made above by the following means: It indicates, for instance, that J-SparseFI-HM produces particularly accurate data in the ultra violet (UV) wavelength range below 400 nm. Remarkable, the CNMF product, which is used to initialize J-SparseFI-HM, performs the worst in this range. This strength of J-SparseFI-HM is particularly useful, if not crucial, for analyses of water, coastal areas and regions with high soil moisture. In contrast, HySure performs better in *exactly* those wavelength ranges in which the multispectral sensor is responsive (see Fig. 57).

ROSIS-3 – *University of Pavia* This data set implicates a comparatively challenging fusion task as it features the highest GSD ratio among all data sets used in this chapter (see Tab. 6), i.e., eight. On the other hand, the limited spectral coverage of the ROSIS-3 hyperspectral instrument mitigates this difficulty to some extend, because no SWIR channels need to be



$\boldsymbol{I}_X$: High-resolution multispectral input image

$\boldsymbol{I}_Y$: Low-resolution hyperspectral input image

$\boldsymbol{I}_{\hat{Z}}$: J-SparseFI-HM fusion result

$\boldsymbol{I}_{Z,\mathrm{ref}}$: High-resolution hyperspectral reference image

Fig. 67. Color composites of the input images (first and second image from the left), the J-SparseFI-HM data fusion product (second image from the right) and the high-resolution hyperspectral reference image (right-most image) corresponding to the CASI *University of Houston* data set. The multispectral input image is shown as true-color composite while all hyperspectral images are shown as false-color composites with 0.75, 0.64, and 0.38 μm being selected for red, green and blue, respectively.

sharpened by VNIR multispectral bands.

Similarly to the color-composite images corresponding to the CASI data set shown and discussed above, Fig. 69 confirms the outstanding quality of the high-resolution hyperspectral data products generated via J-SparseFI-HM. Even with a resolution ratio as high as eight, the fusion product reveals no blurring effects or artifacts like they are observable, for instance, in the experiments on pan-sharpening in Chapter 4.

More tangible information about potential difficulties J-SparseFI-HM and the other fusion algorithms under comparisons have with this data set can be acquired from Fig. 70. As a first observation, the eight SAM maps shown in Fig. 70 (a) indicate relatively large spectral errors in the same specific areas. Looking at Fig. 69, one sees that those areas correspond to *shadows*, which are largely present in the data. Since shaded areas are characterized by very low intensity values in the image and SAM measures the spectral angle of each pixel independent of its intensity, those errors are not noticeable in the color-composite images depicted in Fig. 69 and not as significantly pronounced in the intensity-dependent RMSE maps shown in Fig. 70 (b). What the RMSE maps reveal is that abrupt transitions from shaded (very low intensity) to illuminated (very high intensity) regions cause different levels of difficulties for the different fusion algorithms. Those differences are particularly prominent in the top-center part of the images, which corresponds to the group of buildings shown on the top right of the magnified views in Fig. 69. Clearly observable is that J-SparseFI-HM produces the by far lowest reconstruction errors and, therefore, seems to be the least susceptible to abrupt intensity chances even at a GSD ratio as large as eight.

In Fig. 70 (c), the band-wise evaluation of the PSNR shows that J-SparseFI-HM reconstructs almost every band better than any other algorithm under comparison. Only in a narrow wavelength range around 0.65 μm, which lies exactly in the middle of QuickBird's red

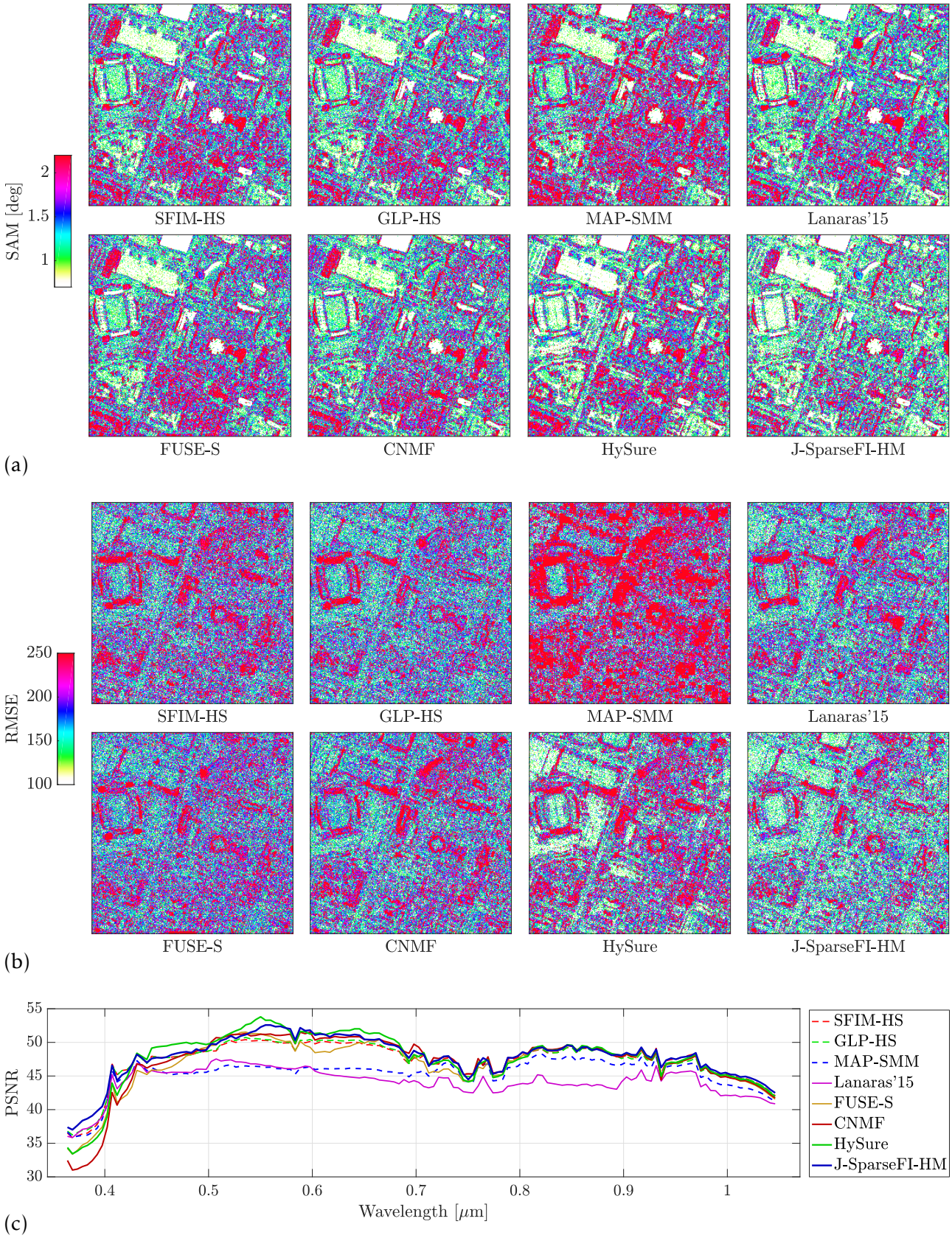
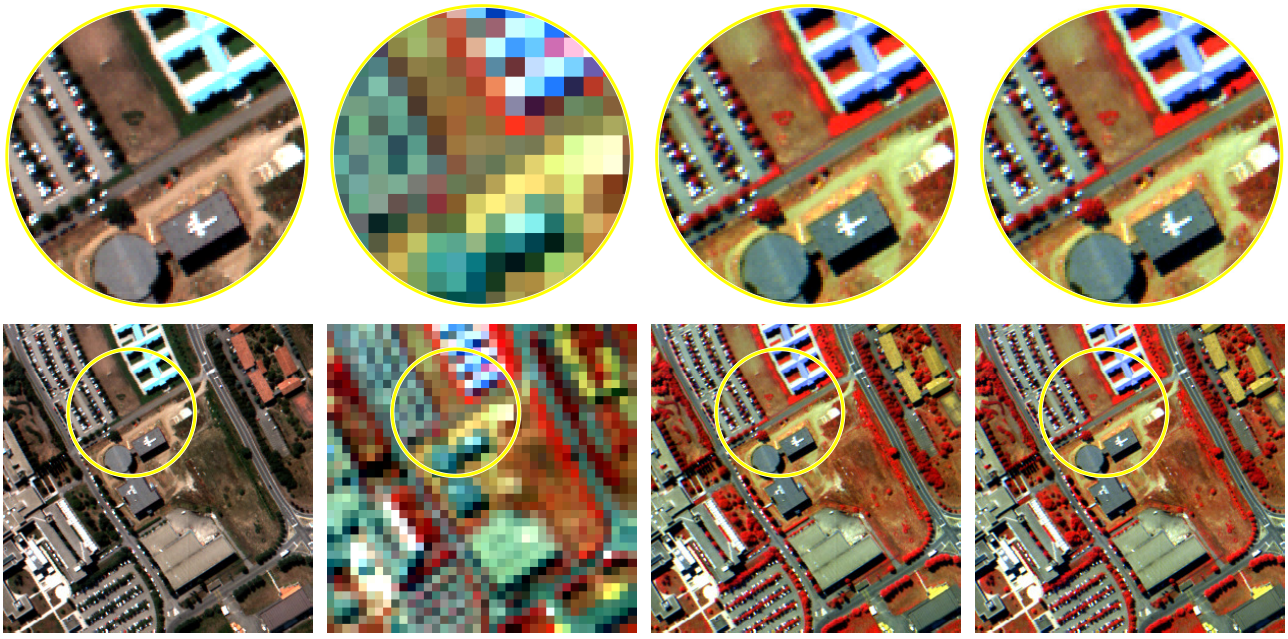


Fig. 68. Qualitative assessment of the data fusion products of all algorithms under comparison based on the CASI *University of Houston* data set. (a) Spectral Angle Mapper (SAM) images indicating the spatial distribution of the spectral deviation of the data fusion products and the high-resolution hyperspectral reference data. (b) Peak Signal-to-Noise Ratio (PSNR) calculated channel-wise indicating the fusion performance over the wavelength range covered by the hyperspectral sensor.



$\boldsymbol{I}_X$: High-resolution multispectral input image | $\boldsymbol{I}_Y$: Low-resolution hyperspectral input image | $\boldsymbol{I}_{\hat{Z}}$: J-SparseFI-HM fusion result | $\boldsymbol{I}_{Z,\mathrm{ref}}$: High-resolution hyperspectral reference image

Fig. 69. Color composites of the input images (first and second image from the left), the J-SparseFI-HM data fusion product (second image from the right) and the high-resolution hyperspectral reference image (right-most image) corresponding to the ROSIS-3 *University of Pavia* data set. The multispectral input image is shown as true-color composite while all hyperspectral images are shown as false-color composites with 0.73, 0.62 and 0.46 μm being selected for red, green and blue, respectively.

channel SRF (see Fig. 57), HySure performs slightly better. In other channels, the PSNR of the HySure product is about average compared to the other methods, which is not in line with the competitive performance HySure achieves with the previously discussed CASI data set or the AVIRIS *Cuprite* data set that is discussed next. One reason for potential problems HySure may have with data sets featuring larger resolution ratios may lie in the PSFs HySure estimates internally. In the cases of lower GSD ratios, in contrast, that estimation procedure seems to have an even positive impact on the fusion results.

The fact that the data fusion product obtained via FUSE-S shows the second best results in this test case confirms the observations made in Chapter 4: Patch-based fusion of images is particularly useful and competitive for data with large resolution ratios.

AVIRIS – *Cuprite* As one of the most interesting geological sites for mineral mapping, the Cuprite mountain range in Nedava is a natural target area for upcoming satellite-based hyperspectral missions. Hence, the conduction of multi-sensor data fusion for hyperspectral resolution enhancement of Cuprite data can be expected to become a hot topic in the near future. Unfortunately, the freely available AVIRIS hyperspectral data set of Cuprite, which is one of the most widely used data sets in the hyperspectral signal processing community, is also the by far most *noisy* among all eight data sets used in the experiments of this chapter. The comparative analysis of the fusion results obtained for this data set primarily serves to investigate how well J-SparseFI-HM is able to cope with noisy data in comparison to the state of the art.

A visual inspection of Fig. 71 leads to very similar conclusions that those drawn from Figs. 67 and 69. J-SparseFI-HM reconstructs the reference data with accuracy that makes it very hard to visually discern the original from the reconstructed high-resolution hyperspectral data.

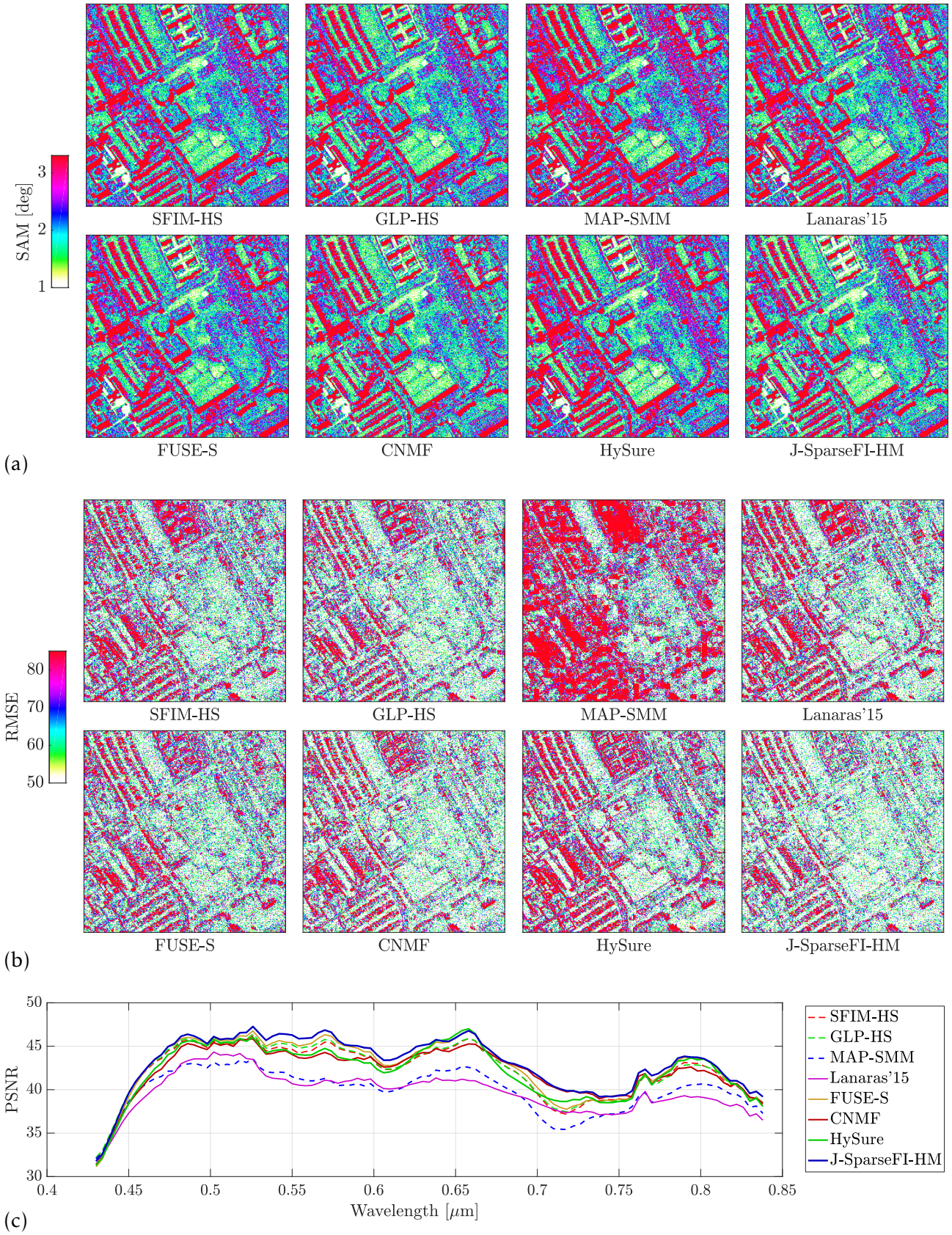


Fig. 70. Qualitative assessment of the data fusion products of all algorithms under comparison based on the ROSIS-3 *University of Pavia* data set. (a) Spectral Angle Mapper (SAM) images indicating the spatial distribution of the spectral deviation of the data fusion products and the high-resolution hyperspectral reference data. (b) Peak Signal-to-Noise Ratio (PSNR) calculated channel-wise indicating the fusion performance over the wavelength range covered by the hyperspectral sensor.

$\boldsymbol{I}_X$: High-resolution multispectral input image

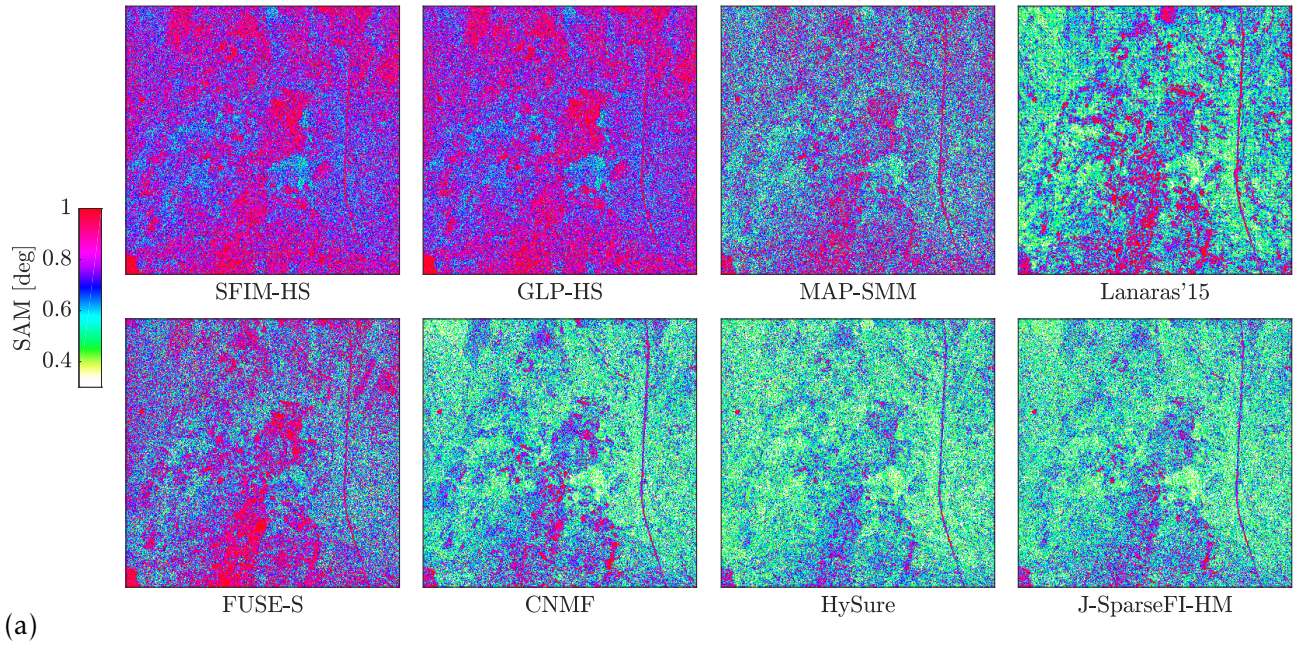
$\boldsymbol{I}_Y$: Low-resolution hyperspectral input image

$\boldsymbol{I}_{\hat{Z}}$: J-SparseFI-HM fusion result

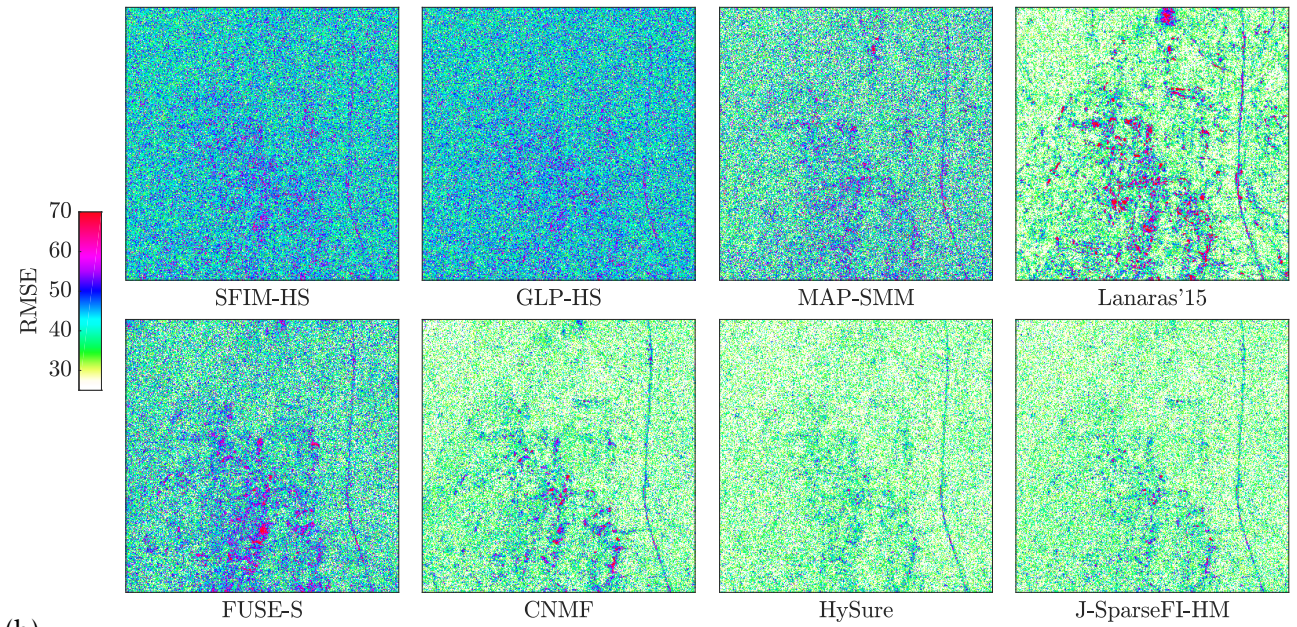
$\boldsymbol{I}_{Z,\text{ref}}$: High-resolution hyperspectral reference image

Fig. 71. Color composites of the input images (first and second image from the left), the J-SparseFI-HM data fusion product (second image from the right) and the high-resolution hyperspectral reference image (right-most image) corresponding to the AVIRIS *Cuprite* data set. The multispectral input image is shown as true-color composite while all hyperspectral images are shown as false-color composites with 2.4, 2.1 and 1.3 μm being selected for red, green and blue, respectively.

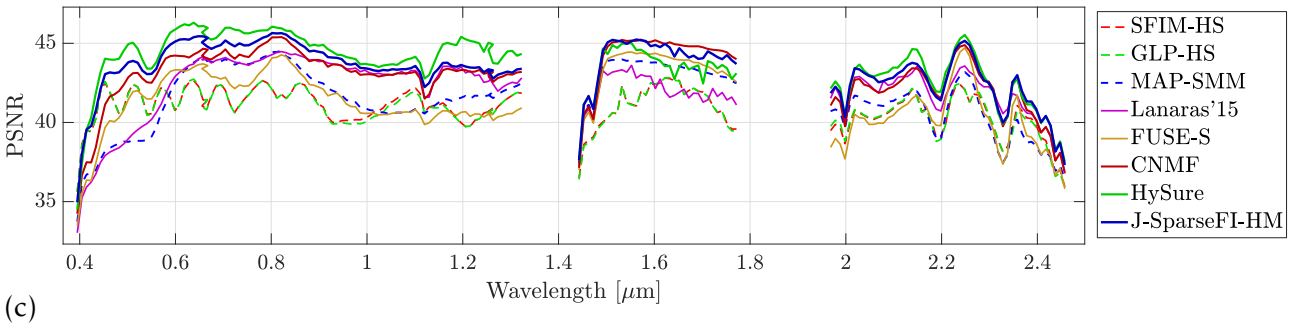
Looking at the error maps shown in Figs. 72 (a) and (b), one can clearly see the high noise level in the data. Noticeable, some of the methods have more difficulties coping with the noise than others. The three fusion products showing, visually, the smallest mean errors in the scene are, first, HySure, second, J-SparseFI-HM and, third, CNMF. Lanaras' methods also performs well over broad areas in the scene but has difficulties locally with distinct objects. This data set is clearly in favor of HySure and Lanaras' method, because they both use TV penalty terms to suppress noise. Therefore, it is no surprise that Lanaras' method shows competitive reconstruction capabilities in relatively homogeneous areas of the scene while struggling locally with heterogeneous features. The denoising property of CNMF is a result reconstructing each pixel as a linear combination of *smooth* spectral profiles produced by VCA. HySure uses both VCA for spectral denoising and TV regularization for spatial denoising, which explains its very competitive performance with this data set. J-SparseFI-HM also denoises in both spectral and spatial domains: In the spatial domain, the three alternating local–non-local–global optimization stages automatically filter out the noise possibly adopted on one of other stages. In the spectral domain, both the robustifying joint sparse recovery of mutually correlation channels in the local–non-local processing module and the SVD-based subspace forward and backward transformations in the global processing module reduce noise. The same ranking follows from analysing the PSNR performance versus the wavelength. Interestingly, HySure performs clearly the best between 0.39 and 1.33 μm. Consistently second best in this wavelength range scores J-SparseFI-HM. CNMF comes third. In the second wavelength range from 1.43 to 1.78 μm, the ranking is exactly opposite. That is, CNMF performs best, closely followed by J-SparseFI-HM and third, with some distance, comes HySure. In the third wavelength range, HySure performs, again, best, then J-SparseFI-HM, and then CNMF. The point of this analysis is that, even though HySure with its TV-based denoising capacity achieves outstanding scores in all *overall* assessment indices, it lacks a little performance in the essential wavelength range between 1.43 to



(a)

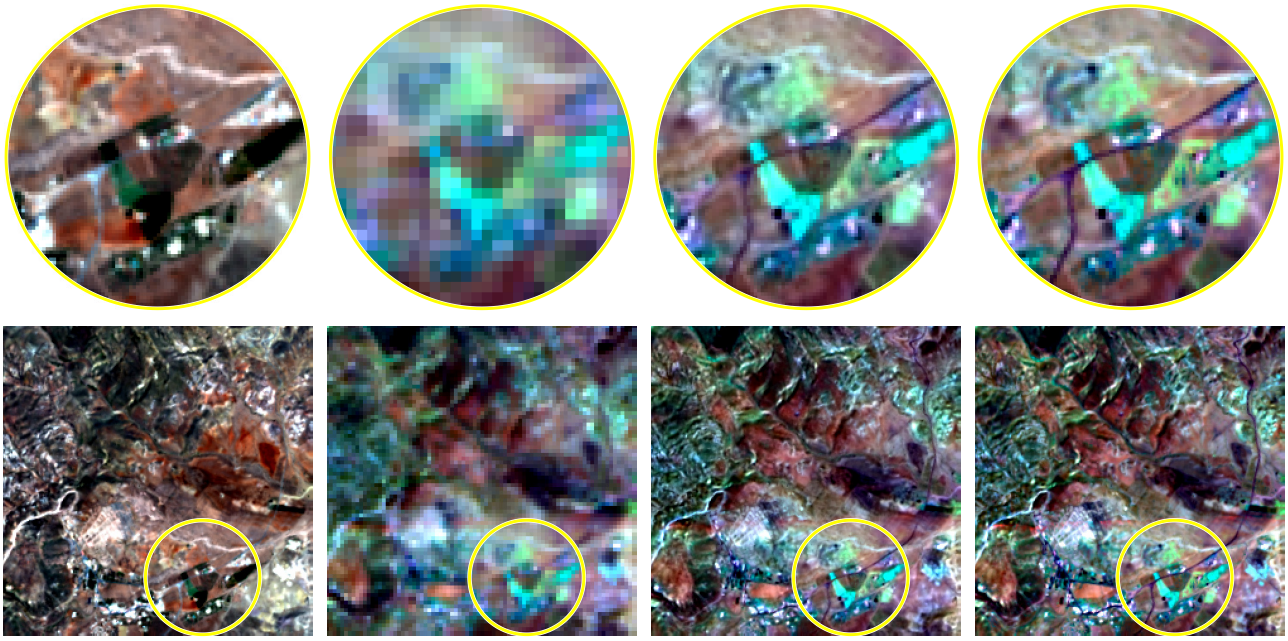


(b)



(c)

Fig. 72. Qualitative assessment of the data fusion products of all algorithms under comparison based on the AVIRIS *Cuprite* data set. (a) Spectral Angle Mapper (SAM) images indicating the spatial distribution of the spectral deviation of the data fusion products and the high-resolution hyperspectral reference data. (b) Peak Signal-to-Noise Ratio (PSNR) calculated channel-wise indicating the fusion performance over the wavelength range covered by the hyperspectral sensor.



$\boldsymbol{I}_X$: High-resolution multispectral input image | $\boldsymbol{I}_Y$: Low-resolution hyperspectral input image | $\boldsymbol{I}_{\hat{Z}}$: J-SparseFI-HM fusion result | $\boldsymbol{I}_{Z,\text{ref}}$: High-resolution hyperspectral reference image

Fig. 73. Color composites of the input images (first and second image from the left), the J-SparseFI-HM data fusion product (second image from the right) and the high-resolution hyperspectral reference image (right-most image) corresponding to the HyMap *Rodalquilar* data set. The multispectral input image is shown as true-color composite while all hyperspectral images are shown as false-color composites with 2.2, 1.1 and 0.75 μm being selected for red, green and blue, respectively.

1.78 μm, which is of highest importance for geologists considering that the absorption features of many minerals lie in exactly that region (see, for instance, the spectral reflectance signature of the Alunite mineral in Fig. 10 on page 16). This analysis is to point out that, depending on the application at hand, the selection of the fusion algorihm best suitable for the conduction of resolution enhancement may require a deeper performance analasis than a simple comparison of averaged assessment scores.

HyMap – *Rodalquilar* This data set is particularly valuable, as it was generated via sophisticate end-to-end simulation procedures that are specifically developed for the generation of realistic EnMAP [Segl et al., 2012] and Sentinel-2 [Segl et al., 2015] products. The simulation was conducted by the authors of the two aforementioned publications who are acknowledged experts on that subject. Moreover, considering that the GSDs of EnMAP and Sentinel-2 are equal to 30 and 10 m, respectively, this data set features the smallest GSD ratio of all eight data sets uses in the experiments of this chapter (see Tab. 6).

First, note that all fusion algorithms under comparison reconstruct the "ground truth", i.e., reference image, reasonably well. This is remarkable, because, for this data set, the reference data was generated *independently* from the input data. With that in mind, the results shown in Figs. 73 and 74 not only confirm the competitiveness of the proposed J-SparseFI-HM fusion algorithm outside standard simulation scenarios, but also validates that the sensor observation model used by the majority of fusion methods under comparison, combined with data-driven estimation of the multispectral sensor's SRFs, approximates the multi-sensor data acquisition process very well.

As for a direct comparison of the assessed quality attained via the different algorithms, the most significant deviation from the results observed for the previous data sets is the performance improvement of Lanaras' method in terms of RMSE and PSNR. One explanation for that might be the possibility of Lanaras' method performing better at lower GSD ratios.

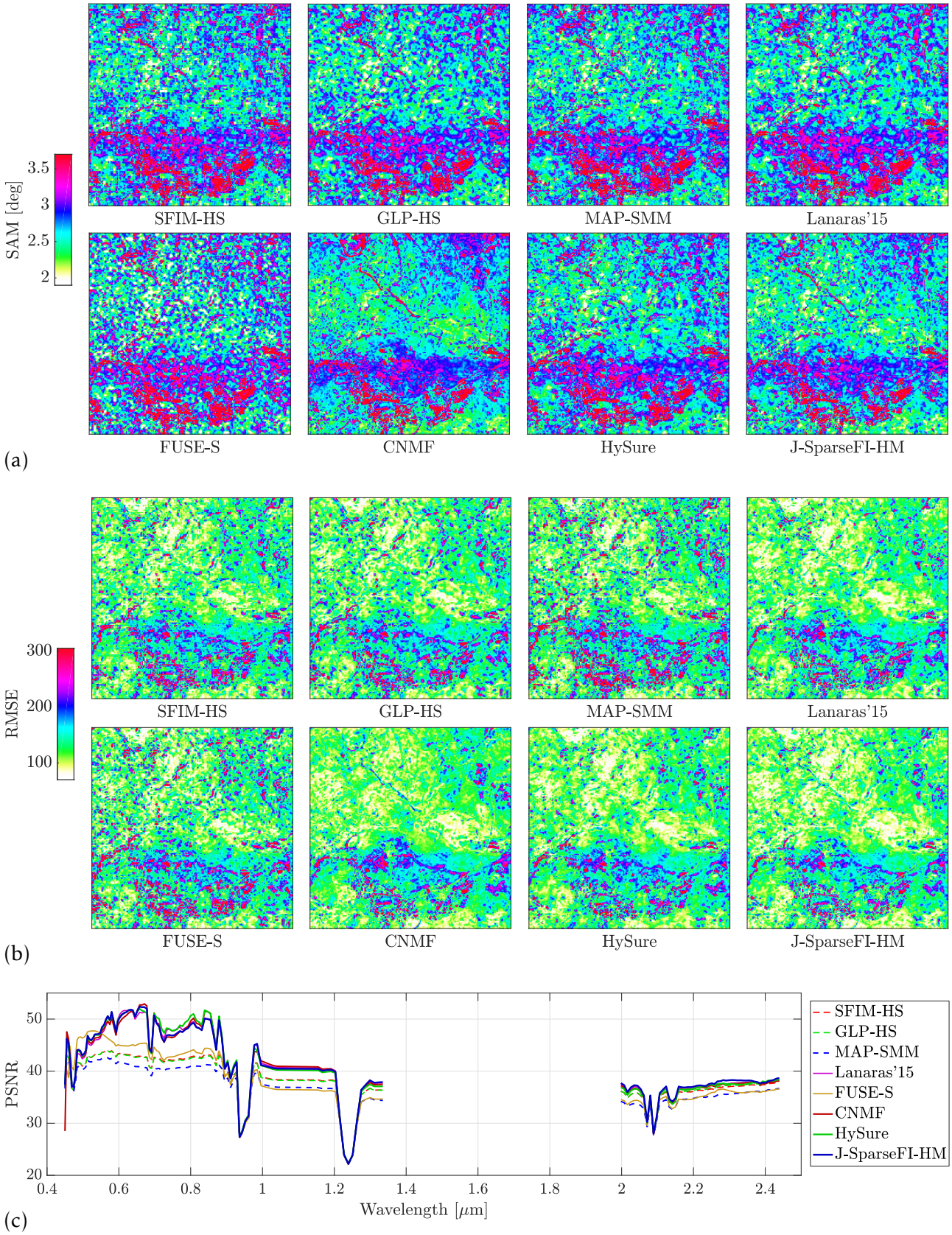
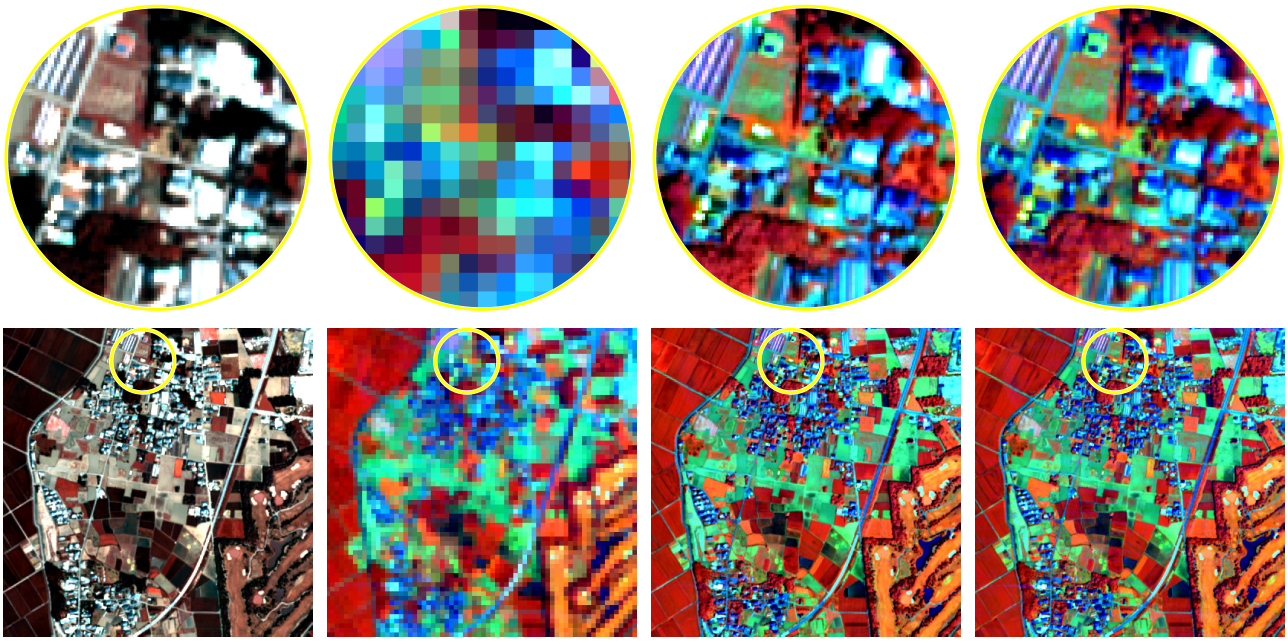
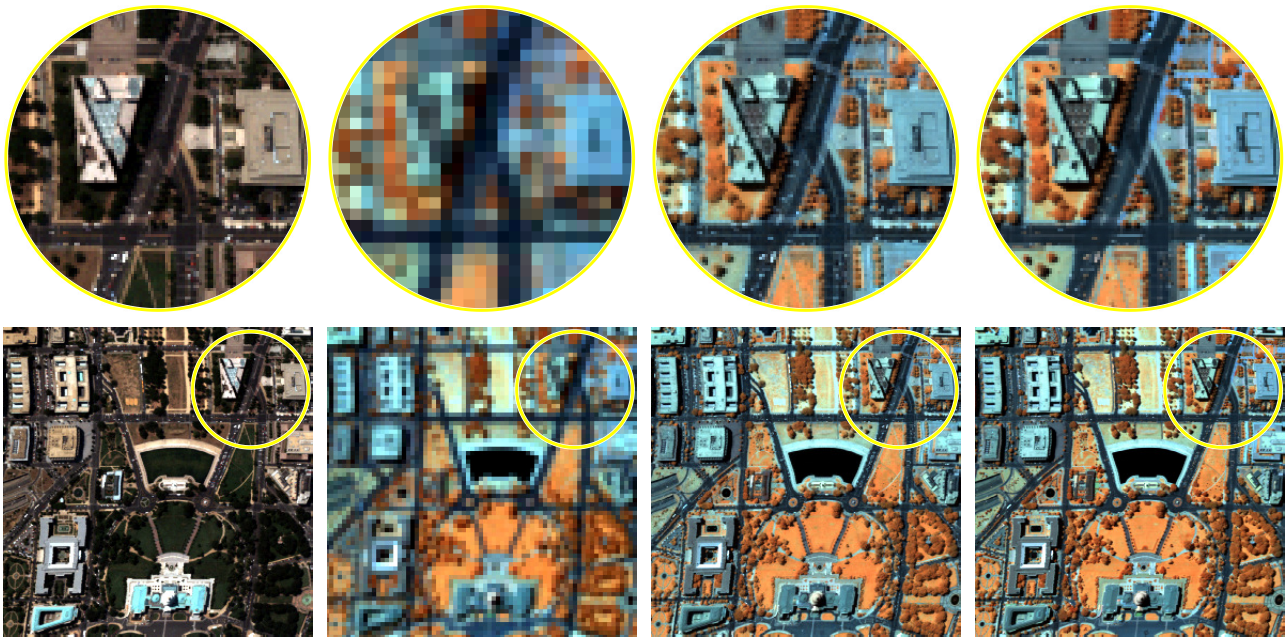


Fig. 74. Qualitative assessment of the data fusion products of all algorithms under comparison based on the HyMap *Rodalquilar* data set. (a) Spectral Angle Mapper (SAM) images indicating the spatial distribution of the spectral deviation of the data fusion products and the high-resolution hyperspectral reference data. (b) Peak Signal-to-Noise Ratio (PSNR) calculated channel-wise indicating the fusion performance over the wavelength range covered by the hyperspectral sensor.



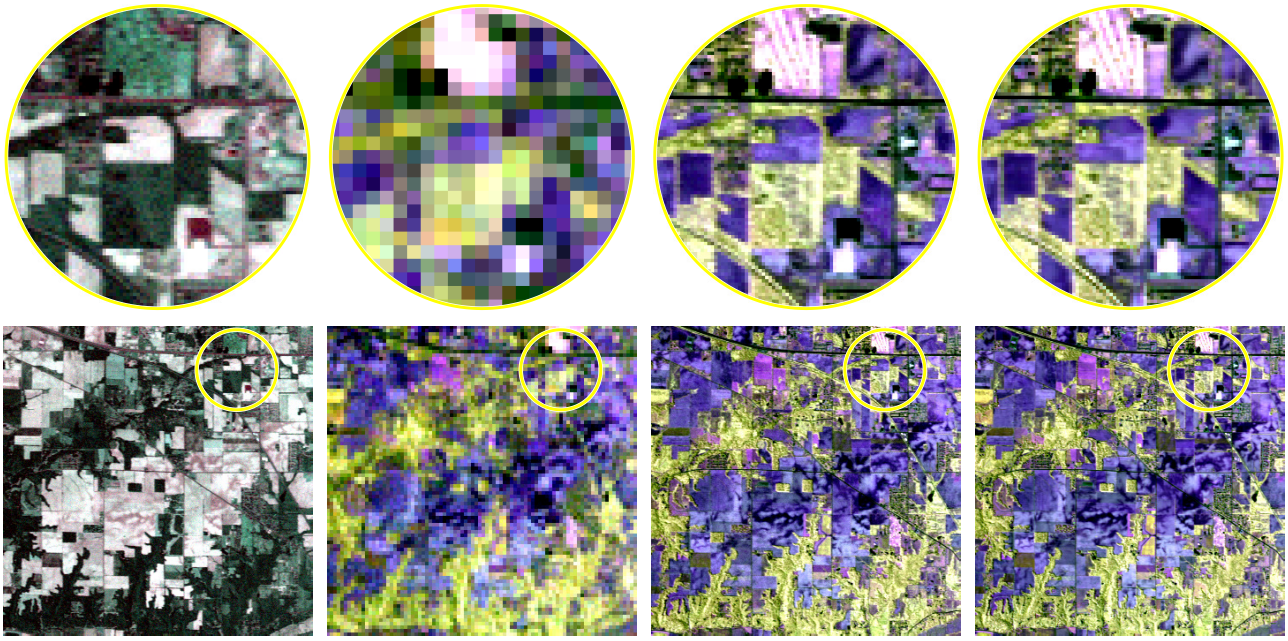
$\boldsymbol{I}_X$: High-resolution multispectral input image | $\boldsymbol{I}_Y$: Low-resolution hyperspectral input image | $\boldsymbol{I}_{\hat{Z}}$: J-SparseFI-HM fusion result | $\boldsymbol{I}_{Z,\mathrm{ref}}$: High-resolution hyperspectral reference image

Fig. 75. Color composites of the input images (first and second image from the left), the J-SparseFI-HM data fusion product (second image from the right) and the high-resolution hyperspectral reference image (right-most image) corresponding to the Hyperspec *Chikusei* data set. The multispectral input image is shown as true-color composite while all hyperspectral images are shown as false-color composites with 0.76, 0.7 and 0.38 μm being selected for red, green and blue, respectively.



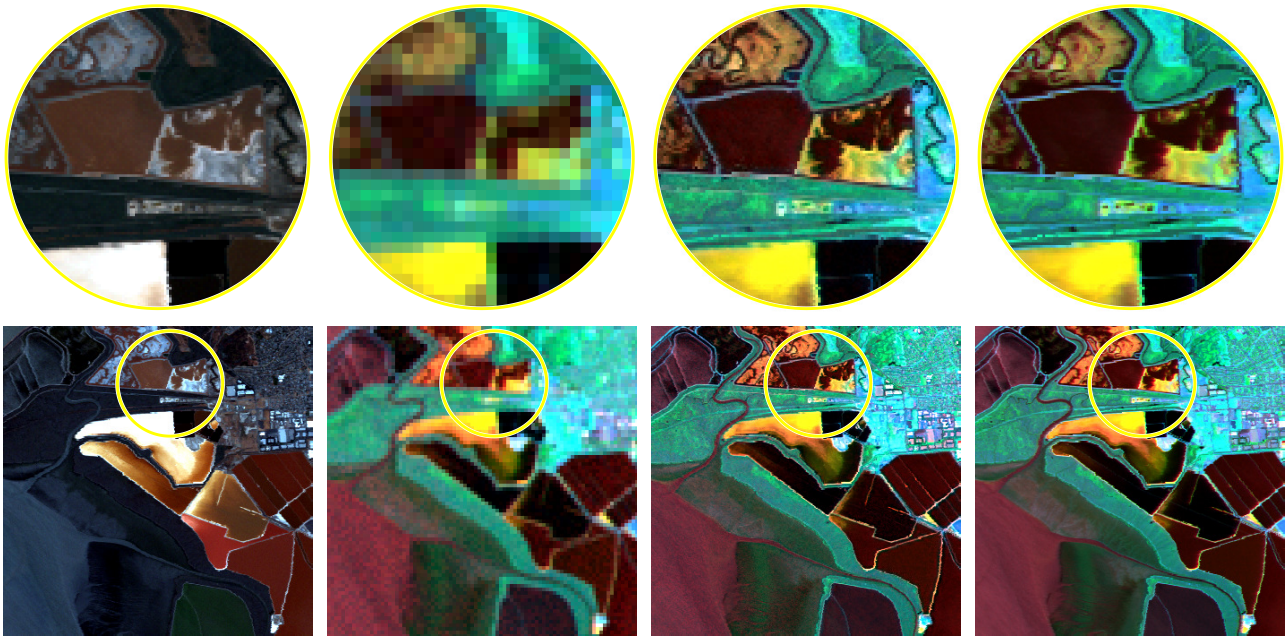
$\boldsymbol{I}_X$: High-resolution multispectral input image | $\boldsymbol{I}_Y$: Low-resolution hyperspectral input image | $\boldsymbol{I}_{\hat{Z}}$: J-SparseFI-HM fusion result | $\boldsymbol{I}_{Z,\mathrm{ref}}$: High-resolution hyperspectral reference image

Fig. 76. Color composites of the input images (first and second image from the left), the J-SparseFI-HM data fusion product (second image from the right) and the high-resolution hyperspectral reference image (right-most image) corresponding to the HYDICE *Washington D.C. Mall* data set. The multispectral input image is shown as true-color composite while all hyperspectral images are shown as false-color composites with 1.1, 1.7, and 2.2 μm being selected for red, green and blue, respectively.



$\boldsymbol{I}_X$: High-resolution multispectral input image | $\boldsymbol{I}_Y$: Low-resolution hyperspectral input image | $\boldsymbol{I}_{\hat{Z}}$: J-SparseFI-HM fusion result | $\boldsymbol{I}_{Z,\mathrm{ref}}$: High-resolution hyperspectral reference image

Fig. 77. Color composites of the input images (first and second image from the left), the J-SparseFI-HM data fusion product (second image from the right) and the high-resolution hyperspectral reference image (right-most image) corresponding to the AVIRIS *Indian Pines* data set. The multispectral input image is shown as true-color composite while all hyperspectral images are shown as false-color composites with 1.1, 0.75 and 1.3 μm being selected for red, green and blue, respectively.



$\boldsymbol{I}_X$: High-resolution multispectral input image | $\boldsymbol{I}_Y$: Low-resolution hyperspectral input image | $\boldsymbol{I}_{\hat{Z}}$: J-SparseFI-HM fusion result | $\boldsymbol{I}_{Z,\mathrm{ref}}$: High-resolution hyperspectral reference image

Fig. 78. Color composites of the input images (first and second image from the left), the J-SparseFI-HM data fusion product (second image from the right) and the high-resolution hyperspectral reference image (right-most image) corresponding to the AVIRIS *Moffett Field* data set. The multispectral input image is shown as true-color composite while all hyperspectral images are shown as false-color composites with 0.38, 1.1 and 2.2 μm being selected for red, green and blue, respectively.

Moreover, the scene is comparatively little heterogeneous (see also Fig. 55), which is in favor of the TV-regularization-based methods HySure and Lanaras's method. Apart from that, the experimental results shown in Figs. 73 (a), (b) and (c) collectively suggest that CNMF, HySure and, as from now on, J-SparseFI-HM lead the field in hyperspectral-multispectral data fusion.

Remaining Data Sets More visual impressions of J-SparseFI-HM data fusion products together with corresponding input and reference data are provided in Figs. 75, 76, 77 and 78 for the remaining four data sets.

5.5.2 Overall Quantitative Evaluation

The overall quality of all of the resolution-enhanced hyperspectral data is further investigated in a quantitative manner. In particular, the fusion products are assessed by using the four quality measures described in Section 5.3.3, i.e., average PSNR, average SAM, ERGAS and $Q2^n$. Note that the $Q2^n$ quality index was not used in the previous section on parameter analyses. This is simply because of the relatively long processing time it takes to compute the $Q2^n$ value for a single image. It was, therefore, not used for the parameter tuning of any of the fusion algorithms, but only applied to the final products.

Tab. 9 shows the overall quality scores measured for the fusion products obtained for data sets 1–3, Tab. 10 for data sets 4–6, and Tab. 11 for data sets 7 and 8. The best results are shown in bold type, and the second-best results are indicated with underlining. As is discussed in Section 5.3.4, and described in detail in [Yokoya et al., 2017], those tables show the best performance each method is able to achieve after individual parameter tuning for each data set.

To obviate doubts in the numbers shown in those tables, the following note should be made: Since the data sets and fusion results of all methods except for J-SparseFI-HM, that are used for comparison in this chapter, are the very same data sets and results that were obtained for and used in the neutral comprehensive review article [Yokoya et al., 2017], those tables as well as the above-discussed error maps and PSNR plots are representative results of a transparent and fair comparative study. All fusion results are reproducible.

From the tables, the following four observations can be made with regard to J-SparseFI-HM:

(1) The data fusion products obtained with J-SparseFI-HM are of great quality for every tested fusion scenario. Its assessment values are best or second best for almost every quality index and data set.
(2) J-SparseFI-HM copes equally well with hyperspectral and multispectral sensor combinations that feature GSD ratios as high as 3 (data set 8), 4 (data sets 4, 5 and 7), 5 (data sets 1 and 6), 6 (data set 2) and 8 (data set 3).
(3) Relative to the other methods under comparison, J-SparseFI-HM performs outstandingly well if the hyperspectral instrument comprises only VNIR bands that are relatively well covered by the corresponding multispectral channels (data sets 1–3). This scenario is of particular interest, as it foreshadows a great potential for the fusion of DESIS data, which will be VNIR channels only (see Tab. 1), and the 10-m GSD channels of Sentinel-2.
(4) With partly significant difference to the second best method, the proposed J-SparseFI-HM masters the much more challenging fusion scenario, in which VNIR *and* SWIR hyperspectral channels are sharpened with only VNIR multispectral high-resolution channels, superiorly (data sets 4, 5 and 8). Obviously, the results obtained

	CASI – *Houston University*				Hyperspec – *Chikusei*				ROSIS-3 – *University of Pavia*			
Method	PSNR	SAM	ERGAS	$Q2^n$	PSNR	SAM	ERGAS	$Q2^n$	PSNR	SAM	ERGAS	$Q2^n$
SFIM-HS	47.25	1.496	1.016	0.9869	46.60	1.319	1.264	0.9193	42.24	2.619	0.764	0.8972
GLP-HS	47.25	1.473	1.017	0.9870	46.91	1.218	1.281	0.9500	42.34	2.609	0.755	0.8945
MAP-SMM	45.32	1.670	1.166	0.9841	43.27	1.314	1.445	0.9133	39.97	2.876	0.918	0.8368
Lanaras'15	44.12	1.463	1.290	0.9848	42.47	1.337	1.702	0.9198	39.73	2.744	0.946	0.7769
FUSE-S	46.80	1.597	1.159	0.9807	46.65	1.425	1.587	0.9519	42.65	2.694	0.776	0.9035
CNMF	47.10	1.526	1.325	0.9750	46.50	1.250	1.504	**0.9548**	42.19	2.580	0.795	0.8734
HySure	47.66	1.444	1.120	0.9835	47.08	1.177	1.508	0.9524	42.20	2.779	0.808	0.8901
J-SparseFI-HM	**47.81**	**1.366**	**0.941**	**0.9896**	**47.73**	**1.158**	**1.257**	0.9537	**43.15**	**2.510**	**0.729**	**0.9104**

Table 9. Overall quantitative quality assessment results measured for the fusion products J-SparseFI-HM and the state of the art in hyperspectral-multispectral data fusion generated for the CASI, *University of Houston*, data set (left), the Hyperspec, *Chikusei*, data set (center) and the ROSIS-3, *University of Pavia*, data set (right). The best results are shown in bold type, and the second-best results are indicated with underlining.

	HYDICE – *Washington, D.C., Mall*				AVIRIS – *Indian Pines*				AVIRIS – *Cuprite*			
Method	PSNR	SAM	ERGAS	$Q2^n$	PSNR	SAM	ERGAS	$Q2^n$	PSNR	SAM	ERGAS	$Q2^n$
SFIM-HS	38.01	1.803	3.234	0.9729	40.74	0.841	0.400	0.9130	40.87	0.798	0.314	0.9702
GLP-HS	38.14	1.766	3.114	0.9745	41.30	0.826	0.375	0.9524	40.82	0.803	0.316	0.9784
MAP-SMM	35.60	2.208	3.783	0.9606	42.74	0.709	0.336	**0.9780**	41.47	0.696	0.307	0.9734
Lanaras'15	37.45	1.788	3.689	0.9708	41.18	0.782	0.412	0.9740	42.11	0.633	0.288	0.9753
FUSE-S	37.87	2.067	3.254	0.9688	40.05	0.829	0.442	0.8661	41.38	0.747	0.310	0.9785
CNMF	38.61	1.706	3.132	0.9747	41.56	0.642	0.362	0.9227	42.98	0.584	0.258	0.9850
HySure	38.13	1.810	3.474	0.9684	42.44	**0.623**	0.328	0.9320	**43.72**	**0.546**	**0.235**	0.9788
J-SparseFI-HM	**39.17**	**1.679**	**2.743**	**0.9793**	**43.35**	0.627	**0.296**	0.9713	43.39	0.564	0.242	**0.9886**

Table 10. Overall quantitative quality assessment results measured for the fusion products J-SparseFI-HM and the state of the art in hyperspectral-multispectral data fusion generated for the HYDICE, *Washington, D.C, Mall*, data set (left), the AVIRIS, *Indian Pines*, data set (center) and the AVIRIS, *Cuprite*, data set (right). The best results are shown in bold type, and the second-best results are indicated with underlining.

	AVIRIS – *Moffett Field*				HyMap – *Rodalquilar*			
Method	PSNR	SAM	ERGAS	$Q2^n$	PSNR	SAM	ERGAS	$Q2^n$
SFIM-HS	34.32	2.107	5.240	0.8483	38.48	2.669	2.081	0.5858
GLP-HS	35.32	1.969	4.938	0.8562	38.49	2.651	2.072	0.5943
MAP-SMM	34.52	2.393	5.759	0.7513	37.18	2.712	2.259	0.5106
Lanaras'15	35.61	1.957	4.957	0.8351	41.09	2.672	1.929	0.6406
FUSE-S	35.20	2.373	4.826	0.7602	38.83	2.691	2.078	0.6013
CNMF	36.76	1.852	4.013	0.8571	41.04	2.711	2.055	0.7056
HySure	36.20	1.758	4.746	0.8572	41.25	2.679	1.908	**0.7087**
J-SparseFI-HM	**37.65**	**1.548**	**3.590**	**0.9067**	**41.27**	**2.650**	**1.901**	0.6895

Table 11. Overall quantitative quality assessment results measured for the fusion products J-SparseFI-HM and the state of the art in hyperspectral-multispectral data fusion generated for the AVIRIS, *Moffett Field*, data set (left) and the HyMap, *Rodalquilar*, data set (right). The best results are shown in bold type, and the second-best results are indicated with underlining.

for dataset number 8, i.e., simulated EnMAP and Sentinel-2 data, are particularly interesting with a view to the upcoming launch of EnMAP and availability of Sentinel-2 data.

Apart from the largely unrivaled assessment results measured for the J-SparseFI-HM fusion products, the following observations can be made with regard to the other algorithms under comparison [Yokoya et al., 2017]:

(1) The methods SFIM and GLP, which were originally designed for pan-sharpening and adopted to the hyperspectral-multispectral data fusion problem via hypersharpening, demonstrate competitive performance. In particular, when the multispectral SRFs

cover the spectral range of the hyperspectral imager in a balanced manner and the number of multispectral bands is small (e.g., data sets 1–3), the hypersharpening-based methods outperform many of the other methods all of which were specifically designed for the hyperspectral-multispectral fusion task.
(2) Several subspace-based methods (CNMF, Lanaras'15, HySure, and MAP-SMM) show good performance for the AVIRIS data sets 5 and 6, outperforming the hypersharpening methods. This is because the estimation accuracy of high-resolution subspace coefficients or abundances increases as the number of multispectral bands increases.
(3) The performance of the unmixing-based methods is good and stable as a whole. In particular, the unmixing-based methods outperform the Bayesian methods when the SRF overlap is low (e.g., data sets 4, 7, and 8).
(4) The unmixing-based methods show their advantage in the spectral quality measured by SAM for all of the data sets.

5.5.3 Impact of Hyperspectral Resolution Enhancement on Classification

Finally, the quality of the fusion results is indirectly validated via pixel-wise classification, which is one of the most relevant topics in the analysis of hyperspectral remote sensing data. The experiments are an extension of those presented in [Yokoya et al., 2017] with classification results obtained for J-SparseFI-HM added. Since the CASI data set of the University of Houston has been widely used for validating classification performance in the hyperspectral image-processing community [Debes et al., 2014] due to the availability of ground-truth information, classification is performed on that data set. The classification performances that were obtained using the fusion products are compared to those obtained using the reference and input hyperspectral and multispectral images. *Support vector machines* (SVMs) [Vapnik, 2000] and *rotation forests* (RoFs) [Rodriguez and Kuncheva, 2006] were used as classifiers because of their good performance for classifying hyperspectral data with a small number of training samples [Melgani and Bruzzone, 2004; Xia et al., 2014]. The class names and numbers of training and test samples are summarized in Tab. 12. The classification performance is quantitatively evaluated using the overall accuracy (OA), the average accuracy (AA), and the kappa coefficient κ. Because training samples were selected randomly, the classification tests were repeated ten times.

Tab. 13 shows the results of the average OA, AA, and κ over all trials. Fig. 79 presents the classification maps of one trial obtained via RoF for the hyperspectral input data (upper right), multispectral input data (lower left), and J-SparseFI-HM fusion product (lower right) together with a test sample map (upper right). From the results, the following four observations can be made:

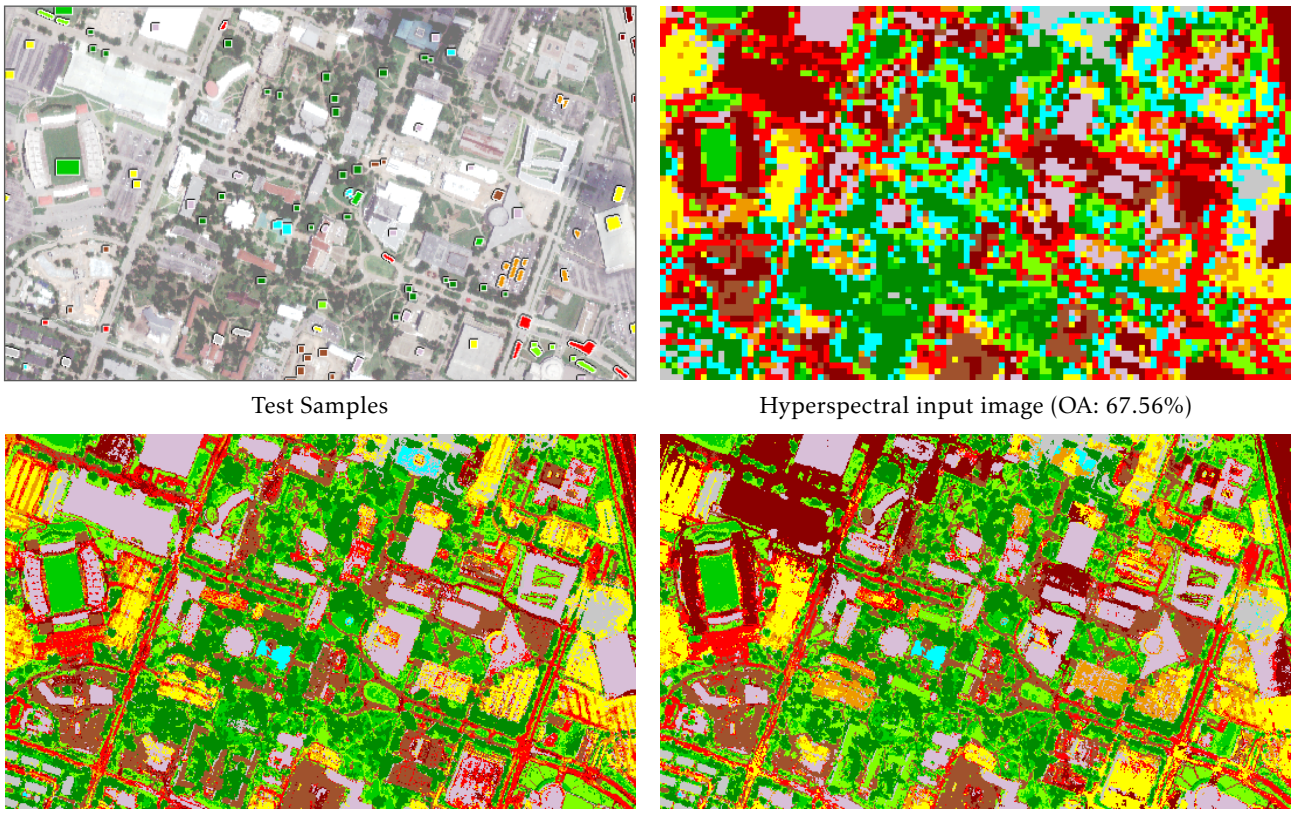
(1) The classification accuracies obtained via RoFs for all of the fusion products are higher than those using the input data. This essential result demonstrates the benefit of hyperspectral-multispectral data fusion for material classification. This is also visually confirmed in Fig. 79. The classification map obtained for the J-SparseFI-HM fusion product shows a much finer level of spatial details than the one obtained for the input hyperspectral image. Although the classification map of the input multispectral image is spatially detailed, the J-SparseFI-HM classification map demonstrates a better material discrimination capability than the classification result obtained for the multispectral image.
(2) The average OA, AA, and κ are partly consistent with the fusion quality assessment results reported in Tab. 9. For instance, J-SparseFI-HM and HySure perform best in the

Class name	# training samples	# test samples	Color used in Fig. 79
Healthy grass	9	384	
Stressed grass	9	199	
Trees	9	507	
Soil	9	227	
Water	9	143	
Residential	9	165	
Commercial	9	308	
Road	9	302	
Highway	9	118	
Parking lot 1	9	365	
Parking lot 2	9	215	

Table 12. Classification results of all fusion products on the CASI, *University of Houston*, dataset.

Classifier	SVM			ROF		
Accuracy	OA	AA	κ	OA	AA	κ
Reference	74.54	72.95	71.51	78.47	77.06	75.86
HS	73.65	72.55	70.56	67.56	66.11	63.84
MS	71.71	70.59	68.40	71.70	70.12	68.33
SFIM-HS	71.96	70.35	68.65	73.29	71.67	70.10
GLP-HS	72.21	70.38	68.89	74.33	72.79	71.26
MAP-SMM	71.09	70.12	67.71	74.14	72.88	71.05
Lanaras'15	75.10	**73.60**	72.15	74.92	73.17	71.93
CNMF	72.27	71.15	69.01	74.57	73.10	71.54
FUSE-S	71.29	69.94	67.90	73.16	71.73	69.96
HySure	**75.18**	73.32	**72.22**	**79.48**	**77.91**	**77.03**
J-SparseFI-HM	74.96	73.46	71.98	77.24	76.14	74.52

Table 13. Quantitative classification results of all fusion products on the CASI, *University of Houston*, dataset.



Test Samples

Hyperspectral input image (OA: 67.56%)

Multispectral input image (OA: 71.70%)

J-SparseFI-HM product (OA: 77.24%)

Fig. 79. Classification maps obtained via rotation forests (RoFs) for the hyperspectral input data (upper right), the multispectral input data (lower left) and the J-SparseFI-HM fusion product (lower right). On the upper left, the test samples of the available ground-truth measurements are overlaid with a brightened true-color composite of the underlying CASI, *University of Houston*, data set.

RoFs-based classification tests and were also found to generate, in average, the most high-quality fusion products in terms of statistical measures in the previous two sections. Moreover, MAP-SMM performs relatively poorly in both statistics-based and classification tests.

(3) In contrast, for some methods, the relative classification performance does not confirm the ranking in the statistical quantitative assessment tests discussed earlier. For example, the pan-sharpening-based methods SFIM-HM and GLP-HM show competitive quality assessment results for that data set in Tab. 9 but less competitive classifica-

tion accuracy results in the present experiment. Furthermore, the classification results calculated for HySure indicate an overall higher classification accuracy than those obtained for the J-SparseFI-HM product. Similarly, Lanaras' method shows a relative high classification accuracy, whereas the statistical measures indicate a rather average quality. The reason for Lanaras' method and, even more so, HySure leading to such high classification accuracies can be well explained by their strong denoising capacity due to TV regularization. Many of the test samples were acquired at large homogeneous places, such as the field in the stadium, parking lots, highways, large planar building roofs, etc. Those areas are exactly where TV further homogenizes the resulting hyperspectral data. If the test samples had been acquired at more isolated objects and heterogeneous ground, the ranking in the classification test would most likely be different.

(4) Some of the fusion products show higher classification accuracies than those calculated for the reference data. This is mainly because of the denoising property of the respective hyperspectral-multispectral data fusion methods.

It should be noted that classification-based validation of hyperspectral-multispectral fusion results does not reflect the absolute quality of fused data – firstly, because the ground-truth information is available only for a comparatively small spatial subset of the data and, secondly, because of the many factors and steps in the data preprocessing chain that may influence the classification capabilities relative to the ground-truth data. The characteristics of applications and data analysis methods highly influence the validation results. It is possible, for instance, to design a fusion method specifically for one classification, application, or data analysis scenario for which it incorporates as much information in its underlying mathematical fusion model as available. Despite its presumably outstanding performance in that scenario, it would certainly fail in a more diverse and comprehensive testing setup, such as the one presented in Sections 5.5.1 and 5.5.2, and, more elaborately, in [Yokoya et al., 2017]. If one is interested in a fusion method for rather simple classification tasks in which the materials of interest are widespread in homogeneous areas, the experiments presented in this section indicate that HySure is likely to be a good choice. For more complex scenarios, in which the important materials can be expected to be found in heterogeneous regions or sparsely distributed in the scene, J-SparseFI-HM is recommended instead, considering the consistently high quality measured for the J-SparseFI-HM products in the comprehensive and divers tests presented in the previous two sections, which implicates a highly versatile scope of application.

6

Conclusion and Perspectives

6.1 Summary and Conclusion

As was motivated in the introduction given in Chapter 1, the inevitable trade-off between the spatial and spectral resolution of optical imaging sensors, additional constraints associated with the large separation between satellite-based remote sensing instruments and the observed surface, and the number and diversity of optical Earth observation missions demand multi-sensor data fusion solutions.

Combining data of different spectral and spatial characteristics to construct products representative of an ideal synthetic sensor is an inherently ill-posed problem. The difficulty arises in maximising both spatial and spectral resolutions, as the greater the difference between the spatial resolutions and number of spectral channels, the greater the number of degrees of freedom.

This dissertation presents sophisticated solutions to two of the most demanding multi-sensor data fusion problems in remote sensing at present, namely pan-sharpening and the fusion of hyperspectral and multispectral data. Two algorithms are thoroughly designed and tailored to the respective fusion problems, i.e., Jointly Sparse Fusion of Images (J-SparseFI) for pan-sharpening and Jointly Sparse Fusion of Hyperspectral and Multispectral Imagery (J-SparseFI-HM) for hyperspectral-multispectral data fusion.

The issues associated with those problems being severely ill-posed are handled by

- ◇ embedding both information about physical properties of the original and synthesized sensing systems and statistical aspects of the processed and generated data into the underlying models describing the inter-sensor relationships,
- ◇ exploiting prior knowledge about image patches featuring sparse representations in suitable dictionaries, to break down the overall fusion problem to a large number of small, regularized convex optimization problems, and
- ◇ the novel idea of utilizing the mutual correlation between multi- and hyperspectral channels to jointly process groups of channels via joint sparse recovery, which is demonstrated to have a robustifying impact on the accuracy of the estimated jointly sparse coefficients.

Specific conclusions that can be drawn from the results achieved with both algorithms that are presented in Chapters 4 and 5 can be summarized as follows.

The J-SparseFI Solution of the Pan-sharpening Problem In Chapter 4, an improved, fully parallelized version of the recently proposed SparseFI algorithm is presented, which incorporates several new features that are elaborately assessed and demonstrated to lead to significant fusion quality improvements and computational speedups. Key features of the J-SparseFI algorithm are listed below:

- The signal structural correlation between multispectral channels is elaborately utilized and framed into a *joint sparsity model* (JSM). The JSM makes use of the distributed compressive sensing theory that restricts the solution of an underdetermined system by considering an ensemble of signals being jointly sparse.
- To overcome the spectral mismatch between the panchromatic and multispectral sensors, a sophisticated strategy is developed incorporating a sensor spectral and channel mutual correlation analysis. Based on this analysis, multispectral channels are automatically assigned to
  - primary groups of jointly processed channels,
  - individually processed channels, and
  - secondary groups of jointly processed channels.
- A strategy is proposed to systematically sharpen those three categories of channels in a decision-driven manner. As a result of the aforementioned sensor spectral and channel mutual correlation analysis, dictionaries are generated from either the panchromatic image or previously reconstructed high-resolution multispectral channels. This spectral adaptation of the dictionary to specific channel groups is one of the main keys to overcoming the problem many pan-sharpening methods have with multispectral channels not being covered by the panchromatic spectral response function.
- The computational burden associated with earlier works on sparse representation-based pan-sharpening, including SparseFI, is mitigated as follows: First, the dimension of each patch recovery problem is made controllable through an adjustable dictionary size, which, in contrast to previous works, is no longer dependent on the dimensions of the input data. Second, a fully parallelized framework is proposed and implemented, which enables very fast processing. The proposed scheme is designed so that each of the possibly overlapping patches are processed independently instead of sequentially as was proposed in the previous works.
- A *locally adjustive* coupled dictionary pair is generated for each patch reconstruction problem. Compared to previous approaches, which use a single global dictionary containing all candidate atoms, locally adjustive dictionaries are shown to implicate both a computational speedup (brought by potentially smaller dictionary sizes) and improvements regarding the quality of the pan-sharpened data (brought by adaptation to local patch information).

The proposed methodology is thoroughly validated in Chapter 4. Detailed descriptions of all adjustment parameters, comprehensive analyses of their impacts on the quality of the data fusion product and processing time, and recommendations of default settings are provided. The quality of the data fusion products obtained via J-SparseFI is elaborately assessed and compared to the state of the art for a variety of test scenarios on data sets both semi-simulated from airborne HySpex data and real satellite-based WorldView-2 data.

Statistical assessment, based on data sets representing mixed urban and vegetated areas, reveal the superior performance of the proposed J-SparseFI algorithm in direct comparisons to both established and recently proposed pan-sharpening methods. Tests cover both the

case of the standard spatial resolution ratio of 4, and the more extreme resolution ratio of 10, in order to test the effectiveness and limits of J-SparesFI in comparison to the state of the art. A particular strength of the proposed algorithm is that it accurately sharpens even the more difficult NIR channels, whose spectral information contents are not even indirectly contained in the input high-resolution panchromatic data. Visualization and statistical information extracted from difference images between the fusion products and reference data indicate that J-SparseFI recovers spatial details of the high-resolution multispectral data more accurately than its predecessors.

Finally, J-SparseFI can be considered as a high-end pan-sharpening algorithm that solves even severely ill-posed multi-sensor data fusion problems associated with large resolution ratios and relatively large numbers of multispectral bands. However, the high quality of J-SparseFI pan-sharpening products entails two drawbacks: (1) Even though the computational burden is reduced methodologically compared to SparseFI and other related algorithms, and a thorough implementation and powerful processing platform – such as the SuperMUC supercomputer the J-SparseFI implementation is optimized for – is demonstrated to enable fast processing, most conventional pan-sharpening methods are simpler to reimplement and computationally less consumptive; (2) The number of parameters is larger than for most conventional methods. This aspect is restrictive though given the recommendations for parameter settings in Chapter 4, which facilitate adjustments to new fusion scenarios.

The J-SparseFI-HM Solution of the Hyperspectral-Multispectral Data Fusion Problem
In comparison to the pan-sharpening setup, the fusion of hyperspectral satellite data with higher-resolution multispectral satellite data incorporates new challenges to overcome while providing new information to account for. On one hand, the significantly lower spatial resolution hyperspectral satellite data, relative to multispectral imagery, makes most pan-sharpening approaches, including J-SparseFI, impractical for resolution enhancement of hyperspectral data, as they operate primarily in the spatial domain. Moreover, many hyperspectral sensors acquire signals in both the VNIR and SWIR wavelength ranges, whereas most multispectral instruments comprise VNIR channels only. Hence, hyperspectral-multispectral data fusion algorithms must be able to reliably transform high-resolution details to wavelength ranges that reveal different surface and material information than what is captured by VNIR channels. On the other hand, the rich spectral information contained in hyperspectral data, as well as multiple complementary multispectral higher-resolution channels offer new opportunities to be ingeniously used and combined with other aspects of the problem to compensate for the aforementioned challenges.

Equipped with experience gained from developing the J-SparseFI pan-sharpening algorithm and motivated by several upcoming hyperspectral satellite missions, including the German Environmental Mapping and Analysis Program (EnMAP), enormous effort has been made in the scope of this dissertation to design a multi-sensor data fusion solution that is capable of reliably enhancing the spatial resolution of hyperspectral satellite imagery.

Following the development and publication of several intermediate algorithmic approaches to hyperspectral-multispectral data fusion of increasing sophistication [Grohnfeldt et al., 2013, 2014, 2015; Grohnfeldt and Zhu, 2015b], the J-SparesFI-HM algorithm presented in Chapter 5 of this dissertation incorporates the best elements of all previously proposed methods as well as new features that further improve the performance and reliability of the spatially enhanced hyperspectral fusion products.

Unique features of the proposed fusion technique can be summarized as follows:

- ◇ One of the novelties of J-SparseFI-HM lies in its alternating three-stage local–non-

local–global processing procedure: Patch-based reconstruction is implemented in one module combining local with non-local information fusion, whereas a global processing module conducts data fusion on the full spatial extend. On all three levels, a sensor observation model incorporating spectral and spatial sensor characteristics is extensively utilized. A first patch estimate is found as the optimum fit of the high-resolution multispectral channels to the local measurements via least squares regression. The resulting patch is further improved by minimizing the residual error to the measurements via estimation of jointly sparse coefficients that correspond to non-locally distributed patches. This step utilizes fast tools developed in the scope of the modern theory of joint sparse signal recovery. Spatially and spectrally adaptive low- and high-resolution coupled dictionaries are generated based on non-local patch information. The global processing module solves a regularized version of the inverse sensor observation model via convex optimization over the full spatial extent. The latter module enhances the spatial consistency across individually processed patches and spectral consistency over groups of jointly processed hyperspectral channels.

- ◇ A sophisticated *Correlation-based HyperSpectral Grouping* (CorHySpeG) concept was developed for and is embedded in the J-SparseFI-HM algorithm to identify, for each patch-reconstruction task individually and automatically, a set of partially overlapping groups of hyperspectral channels, which are particularly suitable for joint processing.
- ◇ Locally high mutual correlation between hyperspectral channels in the same channel group is utilized as prior information for joint sparse signal recovery, which is elaborately demonstrated in Chapter 5 to improve patch reconstruction accuracy, compared to the accuracy of patches that are reconstructed channel by channel individually.
- ◇ In comparison to existing works, such as the methods proposed in [Grohnfeldt et al., 2014; Grohnfeldt and Zhu, 2015b; Wei et al., 2014, 2015a], more information about the system and data is incorporated in the estimation procedure of the sparse coefficients: In [Wei et al., 2014, 2015a], the sparse coding step is conducted solely on the *high-resolution level* using a synthetic dictionary and low-spectral-dimensional estimates of the target image as measurements; In [Grohnfeldt et al., 2013, 2014, 2015; Grohnfeldt and Zhu, 2015b], the estimation process takes place solely on the *low-resolution level* using low-resolution dictionaries and the low-resolution hyperspectral input data as measurement. The proposed J-SparseFI-HM technique uses both of its multi-resolution coupled dictionaries for robust coefficient estimation, while *both* the *high- and low-resolution* input data are used as measurements in addition to the local hyperspectral patch extracted from the previous high-resolution hyperspectral iteration.

Moreover, the following conclusions can be drawn:

- ◇ Versatility and reliable performance of the presented J-SparseFI-HM data fusion algorithm are positively validated based on the various complementary multi-sensor data fusion scenarios compiled in [Yokoya, Grohnfeldt, and Chanussot, 2017].
- ◇ Color-composite images of the input data, J-SparseFI-HM products and reference data, that are depicted for eight different data sets, visually indicate the consistent high quality of J-SparseFI-HM fusion products, even in cases of spatial resolution ratios as large as eight and fusion of SWIR hyperspectral channels with VNIR-only multispectral data.
- ◇ Comparative quantitative quality assessment of the spatially enhanced hyperspectral data produced via J-SparseFI-HM and seven newly established fusion methods confirm the tremendous potential of sparse-representation-based data fusion and the complementarity of the various concepts and modules combined in J-SparseFI-HM. In seven out of eight fusion scenarios, quantitative evaluation of all fusion products yielded that the J-SparseFI-HM products provide higher reconstruction fidelity than those obtained

with any competitor algorithm.

- ◇ An experiment presented in Chapter 5 demonstrates the improvement of hyperspectral-multispectral data fusion on material classification. In that test, the overall classification accuracy achieved with the J-SparseFI-HM fusion product is consistently higher than the accuracies achieved with each of the input data sets, i.e., low-resolution hyperspectral data and higher-resolution multispectral data.
- ◇ Sensitivity analyses and discussions of all variable parameters facilitate adjusting the J-SparseFI-HM algorithm to hyperspectral-multispectral sensor combinations that are not discussed in this dissertation.
- ◇ Finally, J-SparseFI-HM is a high-end algorithm for the fusion of hyperspectral and higher-resolution multispectral data. It has been comprehensively demonstrated that the J-SparseFI-HM high-resolution hyperspectral fusion products are consistently of excellent quality. The price for this performance clearly lies in the relative complexity of the J-SparseFI-HM algorithm, which involves parameter setting and makes – with the current state of technology – a powerful computing system a prerequisite.

6.2 Future Work

End-to-End Processing Chain The work presented in this dissertation sets important milestones towards end-to-end operational fusion of data acquired by current and next-generation remote sensing systems. In order to reach that goal, essential problems remain to be solved. First of all, both J-SparseFI and J-SparseFI-HM require the low- and high-resolution input data to be co-registered and processed up to the same product level. Those are strong assumption, which are associated with a lot of work. Ideally, a processing chain should be designed and implemented, which jointly processes data of a specific pair of sensors (possibly mounted on multiple platforms), over a defined target area, up to the point where the presented algorithms can be embedded as modules.

Optimization for Specific Missions and Sensor Combinations In both Chapter 4 and 5, recommendations for parameter settings are provided. Nonetheless, optimum parameter values may vary from one sensor pair to another. Hence, best product quality can only be achieved after a thorough sensitivity analysis, possibly based on spectrally and spatially degraded versions of sample input data if no reference data is available.

Further Development of the J-SparseFI Algorithm Some of the features that are newly added to J-SparseFI-HM and introduced in Chapter 5 may have a positive impact also on the quality of pan-sharpened data, if added also to J-SparseFI. In particular, the alternating local–non-local–global processing approach can be adapted relatively easily to the pan-sharpening setup.

Furthermore, from the comprehensive experimental results presented in Section 5.5 one can observe that spatial denoising via, e.g., TV-based regularization as is implemented in the HySure data fusion technique, is well-suited for coping with noisy measurements in the scope of multiresolution data fusion. A TV-regularization term could be relatively easily added to the global processing module if added to J-SparseFI.

Further Development of the J-SparseFI-HM Algorithm Similarly to what is described in the previous paragraph, TV-based regularization might be an asset to J-SparseFI-HM if added to the global processing module.

Recent preliminary tests have revealed that the convergence rate achieved with low-quality

initial hyperspectral input data $\boldsymbol{I}_{\tilde{Z}}$ can be significantly improved by modifying the parameter setting in the first iteration as follows: Only the local part of the local–non-local processing module is used. This can be realized, for instance, by setting $N_a = 1$ or $\lambda \to \infty$. Moreover, both μ_X and μ_Y should be set to "large" values in order to reduce the influence of the potentially very inaccurate initial rough estimate. For all subsequent iterations, parameters should be set as recommended in Chapter 5.

Hyperspectral-Multispectral-Panchromatic Data Fusion – Potentials and Limits In this thesis, only the cases of *two* input images of different spectral and spatial resolution are discussed. The fusion of hyperspectral, multispectral and panchromatic imagery is an interesting problem setup, which has been addressed rarely in the literature [Yokoya et al., 2011b]. Since this dissertation presents both a pan-sharpening and a hyperspectral-multispectral data fusion algorithm, it would be interesting to use both methods in combination to solve the hyperspectral-multispectral-panchromatic data fusion problem either in a sequential manner or by developing a stand-alone methodology which is optimized for a mission-specific set of three sensors.

Abbreviations

Abbriviation	Description	Page
1-D	One Dimentional	13
A/D	Analog-to-Digital	13
AA	Absolute Accuracy	160
ADMM	Alternating Direction Method of Multipliers	69
ALI	Advanced Land Imager	38
ARSIS	*Amélioration de la Résolution Spatiale par Injection de Structures*; engl.: spatial resolution enhancement by structure injection	63
AVIRIS	Airborne Visible/Infrared Imaging Spectrometer	28
AWL	Additive Wavelet Luminance	64
AWLP	Additive Wavelet Luminance Proportional	64
BPDN	Basis Pursuit Denoising	56
BRDF	Bidirectional Reflectance Distribution Function	6
BT	Brovey Transformation	63
CASI	Compact Airborne Spectographic Imager	127
CAVIS	Clouds, Aerosols, Vapors, Ice, and Snow	38
CC	Correlation Coefficient	46
CNMF	Coupled Nonnegative Matrix Factorization	66
CorHySpeG	Correlation-based Hyperspectral Grouping	72
CS	Component Substitution	62
dB	Decibel	20
DESIS	DLR Earth Sensing Imaging Spectrometer	38
DEM	Digital Elevation Model	10
DLR	Deutsches Zentrum für Luft- und Raumfahrt	21
DN	Digital Number	6
DSM	Digital Surface Model	10
EeteS	EnMAP end-to-end Simulation	131
EM	Electromagnetic	6

Abbriviation	Description	Page
EnMAP	Environmental Mapping and Analysis Program	10
EO-1	Earth Observation-1	31
Eq.	Equation	9
ERGAS	Erreur Relative Globale Adimensionnelle de Synthèse	46
Fig.	Figure	6
FISTA	Fast Iterative Soft Thresholding Algorithm	56
FOV	Field of View	11
FPA	Focal Plane Array	11
FUSE	Fast Fusion based on Sylvester Equation	67
FUSE-S	Fast Fusion based on Sylvester Equation with Sparse representation-based regularization	67
FWHM	Full Width at Half Maximum	14
GDAL	Geospatial Data Abstraction Library	134
GFOV	Ground-projected Field of View	11
GIFOV	Ground-projected Instantaneous Field of View	11
GLP	Generalized Laplacian Pyramid	133
GLP-HS	see GLP and HS	133
GRD	Ground Resolution Distance or Ground Resolved Distance	11
GS	Gram Schmidt	63
GSA	Gram-Schmidt Adaptive	133
GSD	Ground Sampling Distance	11
HR	High-Resolution	80
HS	Hyperspectral	38
HS	Hypersharpening	133
HYDICE	Hyperspectral Digital Imagery Collection Experiment	128
HySure	Hyperspectral Superresolution	133
i.e.	id est, Latin for “That is”	5
IFOV	Instantaneous Field of View	13
IHS	Intensity-Hue-Saturation	63
IHTA	Iterative Hard Thresholding Algorithm	58
ISTA	Iterative Soft Thresholding Algorithm	58
ITA	Iterative Thresholding Algorithm	57
J-SparseFI	Jointly Sparse Fusion of Images	75
J-SparseFI-HM	Jointly Sparse Fusion of Hyperspectral and Multispectral Imagery	103
JSM	Joint Sparsity Model	75
KL	Karhunen-Loewe	35
km	Kilometer	38
LASSO	Least Absolute Shrinkage and Selection Operator	56
LDCM	Landsat Data Continuity Mission	49
LiDAR	Light Detection And Ranging	10
LMM	Linear Mixing Model	30
LR	Low-Resolution	80
m	Meter	37
MAP	Maximum A Posteriori	66

Abbriviation	Description	Page
MAP-SMM	see MAP and SMM	133
MHz	Megahertz	8
MMSE	Minimum Mean Square Error	67
MNF	Maximum Noise Fraction	35
MPI	Message Passing Interface	134
MRA	Multiresolution Analysis	62
MS	Multispectral	24
MTF	Modulation Transfer Function	15
MWIR	Midwave Infrared	9
NAPC	Noise-Adjusted Principal Components	35
NASA	National Aeronautics and Space Administration	21
NDVI	Normalized Difference Vegetation Index	26
NIR	Near Infrared	42
nm	Nanometer	34
NSP	Null Space Property	54
OA	Overall Accuracy	160
ODL	Online Dictionary Learning	69
OLI	Operational Land Imager	38
OMP	Orthogonal Matching Pursuit	69
p.	Page	6
P-FISTA	Parallel FISTA	59
PAF	Processing and Archiving Facility	22
Pan	Panchromatic	61
PC	Principal Component	36
PCA	Principal Component Analysis	20
PN-TSSC	Two-Step Sparse Coding method with patch Normalization	94
pp.	Pages	7
PSF	Point Spread Function	11
PSNR	Peak Signal-to-Noise Ratio	46
QCBP	Quadratically Constrained Basis Pursuit	55
Ref.	Reference	38
RF	Random Forest	30
RGB	Red Green Blue	19
RIP	Restricted Isometry Property	54
RMSE	Root Mean Square Error	46
ROSIS	Reflective Optics System Imaging Spectrometer	127
SNR	Signal-to-Noise Ratio	11
SRF	Spectral Response Function	11
s.t.	Subject to	44
S2eteS	Sentinel-2 end-to-end Simulation	131
SAM	Spectral Angle Mapper	47
SFIM	Smoothing Filtered-based Intensity Modulation	133
SFIM-HS	see SFIM and HS	133
SGC	Spectral Grouping Concept	69

Abbriviation	Description	Page
SL1MMER	Scale-down by L1 norm Minimization, Model selection, and Estimation Reconstruction	82
SMM	Stochastic Mixing Model	66
SparseFI	Sparse Fusion of Images	79
SSDC	Spectral and Spatial De-Correlation	126
SuperMUC	name of the supercomputer at the Leibniz Supercomputing Centre near Munich; the MUC suffix is borrowed from the Munich airport code.	134
supp	Support	53
SVD	Singular Value Decomposition	35
SVM	Support Vector Machine	29
SVR-MM	Synthetic Variable Ratio Merging Method	94
SWIR	Shortwave Infrared	9
Tab.	Table	37
TV	Total Variation	66
UAV	Unmanned Aerial Vehicle	5
UIQI	Universal Quality Index	46
VCA	Vertex Component Analysis	66
VNIR	Visible to Near Infrared	10

References

Abrams, M., Pieri, D., Realmuto, V., Wright, R., 2013. Using EO-1 Hyperion Data as HyspIRI Preparatory Data Sets for Volcanology Applied to Mt Etna, Italy. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 6 (2): 375–385.

Acito, N., Diani, M., Corsini, G., 2011. Signal-Dependent Noise Modeling and Model Parameter Estimation in Hyperspectral Images. IEEE Transactions on Geoscience and Remote Sensing 49 (8): 2957–2971.

Aguilar, M. Á., Saldaña, M. d. M., Aguilar, F. J., 2014. Generation and Quality Assessment of Stereo-Extracted DSM From GeoEye-1 and WorldView-2 Imagery. IEEE Transactions on Geoscience and Remote Sensing 52 (2): 1259–1271.

Aiazzi, B., Alparone, L., Argenti, F., Baronti, S., 1999. Wavelet and pyramid techniques for multisensor data fusion: A performance comparison varying with scale ratios. 251–262.

Aiazzi, B., Alparone, L., Baronti, S., Carlà, R., Garzelli, A., Santurri, L., 2017. Sensitivity of Pansharpening Methods to Temporal and Instrumental Changes Between Multispectral and Panchromatic Data Sets. IEEE Transactions on Geoscience and Remote Sensing 55 (1): 308–319.

Aiazzi, B., Alparone, L., Baronti, S., Garzelli, A., 2002. Context-driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis. IEEE Transactions on Geoscience and Remote Sensing 40 (10): 2300–2312.

Aiazzi, B., Alparone, L., Baronti, S., Garzelli, A., Selva, M., 2006. MTF-tailored Multiscale Fusion of High-resolution MS and Pan Imagery. Photogrammetric Engineering & Remote Sensing 72 (5): 591–596.

Aiazzi, B., Baronti, S., Lotti, F., Selva, M., 2009. A Comparison Between Global and Context-Adaptive Pansharpening of Multispectral Images. IEEE Geoscience and Remote Sensing Letters 6 (2): 302–306.

Aiazzi, B., Baronti, S., Selva, M., 2007. Improving Component Substitution Pansharpening Through Multivariate Regression of MS +Pan Data. IEEE Transactions on Geoscience and Remote Sensing 45 (10): 3230–3239.

Akhtar, N., Shafait, F., Mian, A., 2014. Sparse Spatio-spectral Representation for Hyperspectral Image Super-resolution. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (Hrsg.), Computer Vision – ECCV 2014, Lecture Notes in Computer Science, Springer International Publishing, 63–78.

Alganci, U., Sertel, E., Kaya, S., BerkUstundag, B., 2013. A research on agricultural mapping capabilities of the SPOT 6 satellite images. In: 2013 Second International Conference on Agro-Geoinformatics (Agro-Geoinformatics), 93–96.

Alparone, L., Aiazzi, B., Baronti, S., Garzelli, A., 2015. Remote Sensing Image Fusion. CRC Press.

Alparone, L., Aiazzi, B., Baronti, S., Garzelli, A., Nencini, F., Selva, M., 2008. Multispectral and Panchromatic Data Fusion Assessment Without Reference. Photogrammetric Engineering & Remote Sensing 74 (2): 193–200.

Alparone, L., Baronti, S., Garzelli, A., Nencini, F., 2004. A global quality measurement of pan-sharpened multispectral imagery. IEEE Geoscience and Remote Sensing Letters 1 (4): 313–317.

Alparone, L., Wald, L., Chanussot, J., Thomas, C., Gamba, P., Bruce, L. M., 2007. Comparison of Pansharpening Algorithms: Outcome of the 2006 GRS-S Data-Fusion Contest. IEEE Transactions on Geoscience and Remote Sensing 45 (10): 3012–3021.

Amro, I., Mateos, J., Vega, M., Molina, R., Katsaggelos, A. K., 2011. A survey of classical methods and new trends in pansharpening of multispectral images. EURASIP Journal on Advances in Signal Processing 2011 (1): 79.

Andrews, R. W., Pollard, A., Pearce, J. M., 2015. Photovoltaic System Performance Enhancement With Nontracking Planar Concentrators: Experimental Results and Bidirectional Reflectance Function (BDRF)-Based Modeling. IEEE Journal of Photovoltaics 5 (6): 1626–1635.

Aptoula, E., Mura, M. D., Lefèvre, S., 2016. Vector Attribute Profiles for Hyperspectral Image Classification. IEEE Transactions on Geoscience and Remote Sensing 54 (6): 3208–3220.

Asner, G., 2008. Hyperspectral Remote Sensing of Canopy Chemistry, Physiology, and Biodiversity in Tropical Rainforests. In: Hyperspectral Remote Sensing of Tropical and Sub-Tropical Forests. CRC Press, 261–296.

Aspinall, R. J., 2002. Use of logistic regression for validation of maps of the spatial distribution of vegetation species derived from high spatial resolution hyperspectral remotely sensed data. Ecological Modelling 157 (2–3): 301–312.

Aspinall, R. J., Marcus, W. A., Boardman, J. W., 2002. Considerations in collecting, processing, and analysing high spatial resolution hyperspectral data for environmental investigations. Journal of Geographical Systems 4 (1): 15–29.

Avbelj, J., 2015. Fusion of Hyperspectral Images and Digital Surface Models for Urban Object Extraction. Dissertation, Technische Universität München, München.

Bachmann, M., 2007. Automatisierte Ableitung von Bodenbedeckungsgraden durch MESMA-Entmischung. thesis, Uni Würzburg.

Baillarin, S. J., Meygret, A., Dechoz, C., Petrucci, B., Lacherade, S., Tremas, T., Isola, C., Martimort, P., Spoto, F., 2012. Sentinel-2 level 1 products and image processing performances. In: 2012 IEEE International Geoscience and Remote Sensing Symposium, 7003–7006.

Bar, D. E., Raboy, S., 2013. Moving Car Detection and Spectral Restoration in a Single Satellite WorldView-2 Imagery. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 6 (5): 2077–2087.

Bar, D. E., Wolowelsky, K., Swirski, Y., Figov, Z., Michaeli, A., Vaynzof, Y., Abramovitz, Y., Ben-Dov, A., Yaron, O., Weizman, L., Adar, R., 2010. Target Detection and Verification via Airborne Hyperspectral and High-Resolution Imagery Processing and Fusion. IEEE Sensors Journal 10 (3): 707–711.

Baraniuk, R. G., Cevher, V., Duarte, M. F., Hegde, C., 2010. Model-Based Compressive Sensing. IEEE Transactions on Information Theory 56 (4): 1982–2001.

Barnsley, M. J., Settle, J. J., Cutter, M. A., Lobb, D. R., Teston, F., 2004. The PROBA/CHRIS mission: A low-cost smallsat for hyperspectral multiangle observations of the Earth surface and atmosphere. IEEE Transactions on Geoscience and Remote Sensing 42 (7): 1512–1520.

Baron, D., Duarte, M. F., Wakin, M. B., Sarvotham, S., Baraniuk, R. G., 2009. Distributed Compressive Sensing. arXiv:0901.3403 [cs, math] (unpublished) : 1–42.

Basaeed, E., Bhaskar, H., Al-Mualla, M., 2013. Comparative analysis of pan-sharpening techniques on DubaiSat-1 images. In: Proceedings of the 16th International Conference on Information Fusion, Istanbul, Turkey, 227–234.

Başkurt, D. Ö., Ömrüuzun, F., Çetin, Y. Y., 2015. Hyperspectral unmixing based analysis of forested areas. In: 2015 23nd Signal Processing and Communications Applications Conference (SIU), 2329–2332.

Baumgardner, M. F., Biehl, L. L., Landgrebe, D. A., 2015. 220 Band AVIRIS Hyperspectral Image Data Set: June 12, 1992 Indian Pine Test Site 3 .

Beck, A., Teboulle, M., 2009. A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. SIAM Journal on Imaging Sciences 2 (1): 183–202.

Bedini, E., der Meer, F. V., Ruitenbeek, F. V., 2009. Use of HyMap imaging spectrometer data to map mineralogy in the Rodalquilar caldera, southeast Spain. Int. J. Remote Sens. 30 (2): 327–348.

Bellisola, G., Sorio, C., 2011. Infrared spectroscopy and microscopy in cancer research and diagnosis. American Journal of Cancer Research 2 (1): 1–21.

Benner, P., Novaković, V., Plaza, A., Quintana-Ortí, E. S., Remón, A., 2015. Fast and Reliable Noise Estimation for Hyperspectral Subspace Identification. IEEE Geoscience and Remote Sensing Letters 12 (6): 1199–1203.

Berné, O., Helens, A., Pilleri, P., Joblin, C., 2010. Non-negative matrix factorization pansharpening of hyperspectral data: An application to mid-infrared astronomy. In: 2010 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, 1–4.

Bieniarz, J., 2015. Sparse Methods for Hyperspectral Unmixing and Image Fusion. thesis, University of Osnabrück.

Bieniarz, J., Aguilera, E., Zhu, X., Muller, R., Reinartz, P., 2015. Joint Sparsity Model for Multilook Hyperspectral Image Unmixing. IEEE Geoscience and Remote Sensing Letters 12 (4): 696–700.

Bieniarz, J., Cerra, D., Avbelj, J., Reinartz, P., Müller, R., 2011. Hyperspectral Image Resolution Enhancement Based on Spectral Unmixing and Information Fusion. In: ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVIII-4-W19, Copernicus GmbH, Hannover, Germany, 33–37.

Bioucas-Dias, J. M., Nascimento, J. M. P., 2008. Hyperspectral Subspace Identification. IEEE Transactions on Geoscience and Remote Sensing 46 (8): 2435–2445.

Blumensath, T., 2012. Accelerated iterative hard thresholding. Signal Processing 92 (3): 752–756.

Blumensath, T., Davies, M. E., 2009. Iterative hard thresholding for compressed sensing. Applied and Computational Harmonic Analysis 27 (3): 265–274.

Boccardo, P., Mondino, E. B., Tonolo, F. G., 2004. Applications based on orthocorrected high resolution and hyperspectral images. In: Proceedings of ISPRS Congress, Istanbul, Turkey.

Boche, H., Calderbank, R., Kutyniok, G., Vybíral, J. (Hrsg.), 2015. Compressed Sensing and Its Applications. Applied and Numerical Harmonic Analysis. Springer International Publishing, Cham.

Bonis, R. D., Laneve, G., Palombo, A., Pascucci, S., Pignatti, S., Santini, F., Ananasso, C., 2015. The potential impact of the next hyperspectral prisma mission on the natural and anthropogenic hazards management. In: 2015 IEEE 15th International Conference on Environment and Electrical Engineering (EEEIC), 1643–1646.

Boser, B. E., Guyon, I. M., Vapnik, V. N., 1992. A Training Algorithm for Optimal Margin Classifiers. In: Proceedings of the Fifth Annual Workshop on Computational Learning Theory, COLT '92, ACM, New York, NY, USA, 144–152.

Bovolo, F., Bruzzone, L., Capobianco, L., Garzelli, A., Marchesi, S., Nencini, F., 2010. Analysis of the Effects of Pansharpening in Change Detection on VHR Images. IEEE Geoscience and Remote Sensing Letters 7 (1): 53–57.

Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., 2011. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. Foundations and Trends® in Machine Learning 3 (1): 1–122.

Bredies, K., Lorenz, D. A., 2008. Linear Convergence of Iterative Soft-Thresholding. Journal of Fourier Analysis and Applications 14 (5-6): 813–837.

Bullock, K., 2016. WorldView-4: A new satellite and much more. DigitalGlobe Inc., Blog Archive,

August 29, 2016, Retrieved on November 17, 2016 .

Calderón, R., Navas-Cortés, J. A., Lucena, C., Zarco-Tejada, P. J., 2013. High-resolution airborne hyperspectral and thermal imagery for early detection of Verticillium wilt of olive using fluorescence, temperature and narrow-band spectral indices. Remote Sensing of Environment 139: 231–245.

Candela, L., Formaro, R., Guarini, R., Loizzo, R., Longo, F., Varacalli, G., 2016. The PRISMA mission. In: 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 253–256.

Candès, E. J., 2008. The restricted isometry property and its implications for compressed sensing. Comptes Rendus Mathematique 346 (9): 589–592.

Candès, E. J., Romberg, J., Tao, T., 2006a. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. IEEE Transactions on Information Theory 52 (2): 489–509.

Candès, E. J., Romberg, J. K., Tao, T., 2006b. Stable signal recovery from incomplete and inaccurate measurements. Communications on Pure and Applied Mathematics 59 (8): 1207–1223.

Candès, E. J., Tao, T., 2005. Decoding by linear programming. IEEE Transactions on Information Theory 51 (12): 4203–4215.

Capobianco, L., Garzelli, A., Nencini, F., Alparone, L., Baronti, S., 2007. Spatial enhancement of hyperion hyperspectral data through ALI panchromatic image. In: 2007 IEEE International Geoscience and Remote Sensing Symposium, 5158–5161.

Carper, W. J., Lillesand, T. M., Kiefer, R. W., 1990. The use of intensity-hue-saturation transformations for merging SPOT panchromatic and multispectral image data. Photogrammetric Engineering And Remote Sensing 56 (4): 459–467.

Celik, Y. B., Sertel, E., Ustundag, B. B., 2015. Identification of corn and cotton fields using multitemporal Spot6 NDVI data. In: 2015 Fourth International Conference on Agro-Geoinformatics (Agro-Geoinformatics), 23–26.

Chan, J. C. W., Ma, J., de Voorde, T. V., Canters, F., 2011. Preliminary Results of Superresolution-Enhanced Angular Hyperspectral (CHRIS/Proba) Images for Land-Cover Classification. IEEE Geoscience and Remote Sensing Letters 8 (6): 1011–1015.

Chang, C.-I., 2003. Hyperspectral Imaging: Techniques for Spectral Detection and Classification. Kluwer Academic/Plenum Publishers, New York [u.a.].

Chavez, P. S. J., Sides, S., Anderson, J., 1991. Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic. Photogrammetric Engineering and Remote Sensing 57 (3): 295–303.

Chen, J., Huo, X., 2006. Theoretical Results on Sparse Representations of Multiple-Measurement Vectors. IEEE Transactions on Signal Processing 54 (12): 4634–4643.

Chen, Z., Pu, H., Wang, B., Jiang, G.-M., 2014. Fusion of Hyperspectral and Multispectral Images: A Novel Framework Based on Generalization of Pan-Sharpening Methods. IEEE Geoscience and Remote Sensing Letters 11 (8): 1418–1422.

Cheng, G., Zhou, P., Han, J., 2016. Learning Rotation-Invariant Convolutional Neural Networks for Object Detection in VHR Optical Remote Sensing Images. IEEE Transactions on Geoscience and Remote Sensing 54 (12): 7405–7415.

Choi, M., Kim, R. Y., Nam, M.-R., Kim, H. O., 2005. Fusion of multispectral and panchromatic Satellite images using the curvelet transform. IEEE Geoscience and Remote Sensing Letters 2 (2): 136–140.

Chuvieco, E., Huete, A., 2009. Fundamentals of Satellite Remote Sensing, har/com Edition. Routledge Chapman & Hall, Boca Raton, FL.

Cocks, T., Jenssen, R., Stewart, A., Wilson, I., Shields, T., 1998. The HyMap Airborne Hyperspectral Sensor: The System, Calibration and Performance. In: Proc. 1st EARSeL Workshop on Imaging Spectroscopy, Zurich, Switzerland.

Combettes, P., Wajs, V., 2005. Signal Recovery by Proximal Forward-Backward Splitting. Multiscale Modeling & Simulation 4 (4): 1168–1200.

Condessa, F., Bioucas-Dias, J., Kovačević, J., 2016. Supervised Hyperspectral Image Classification With Rejection. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 9 (6): 2321–2332.

Cucci, C., Delaney, J. K., Picollo, M., 2016. Reflectance Hyperspectral Imaging for Investigation of Works of Art: Old Master Paintings and Illuminated Manuscripts. Accounts of Chemical Research 49 (10): 2070–2079.

Dahlke, S., Fornasier, M., Raasch, T., 2012. Multilevel preconditioning and adaptive sparse solution of inverse problems. Mathematics of Computation 81 (277): 419–446.

Dai, Q., Sun, D.-W., Xiong, Z., Cheng, J.-H., Zeng, X.-A., 2014. Recent Advances in Data Mining Techniques and Their Applications in Hyperspectral Image Processing for the Food Industry. Comprehensive Reviews in Food Science and Food Safety 13 (5): 891–905.

Dalponte, M., Bruzzone, L., Gianelle, D., 2012. Tree species classification in the Southern Alps based on the fusion of very high geometrical resolution multispectral/hyperspectral images and LiDAR data. Remote Sensing of Environment 123: 258–270.

d'Angelo, P., Kuschk, G., Reinartz, P., 2014. Evaluation of Skybox Video and Still Image products. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XL-1: 95–99.

Datt, B., McVicar, T. R., Niel, T. G. V., Jupp, D. L. B., Pearlman, J. S., 2003. Preprocessing EO-1 Hyperion hyperspectral data to support the application of agricultural indexes. IEEE Transactions on Geoscience and Remote Sensing 41 (6): 1246–1259.

Daubechies, I., Defrise, M., De Mol, C., 2004. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. Communications on Pure and Applied Mathematics 57 (11): 1413–1457.

Davies, A. G., Chien, S., Baker, V., Doggett, T., Dohm, J., Greeley, R., Ip, F., Castano, R., Cichy, B., Rabideau, G., Tran, D., Sherwood, R., 2006. Monitoring active volcanism with the Autonomous Sciencecraft Experiment on EO-1. Remote Sensing of Environment 101 (4): 427–446.

Davies, M. E., Eldar, Y. C., 2012. Rank Awareness in Joint Sparse Recovery. IEEE Transactions on Information Theory 58 (2): 1135–1146.

de la Fuente, D., Suarez, J., Yagüe, J., Pedrazzani, D., 2013. Potentiality of World-View 2 data for precision agriculture. In: 2013 IEEE International Geoscience and Remote Sensing Symposium - IGARSS, 2825–2828.

Debes, C., Merentitis, A., Heremans, R., Hahn, J., Frangiadakis, N., van Kasteren, T., Liao, W., Bellens, R., Pižurica, A., Gautama, S., Philips, W., Prasad, S., Du, Q., Pacifici, F., 2014. Hyperspectral and LiDAR Data Fusion: Outcome of the 2013 GRSS Data Fusion Contest. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7 (6): 2405–2418.

DigitalGlobe, I., 2016. Digital Globe, Inc. - Satellite Information. https://www.digitalglobe.com/resources/satellite-information, retrieved on October 17, 2016.

Dill, T. J., Rozin, M. J., Palani, S., Tao, A. R., 2016. Colloidal Nanoantennas for Hyperspectral Chemical Mapping. ACS Nano 10 (8): 7523–7531.

Donoho, D. L., 2006. Compressed sensing. IEEE Transactions on Information Theory 52 (4): 1289–1306.

Dou, W., Chen, Y., Li, X., Sui, D. Z., 2007. A general framework for component substitution image fusion: An implementation using the fast image fusion method. Computers & Geosciences 33 (2): 219–228.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. Remote Sensing of Environment 120: 25–36.

Du, Q., Younan, N. H., King, R., Shah, V. P., 2007. On the Performance Evaluation of Pan-Sharpening Techniques. IEEE Geoscience and Remote Sensing Letters 4 (4): 518–522.

Eckardt, A., Horack, J., Lehmann, F., Krutz, D., Drescher, J., Whorton, M., Soutullo, M., 2015. DESIS (DLR Earth Sensing Imaging Spectrometer for the ISS-MUSES platform). In: 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 1457–1459.

Eckert, S., Kneubühler, M., 2004. Application of HYPERION data to agricultural land classification and vegetation properties estimation in Switzerland. In: Proceedings of the ISPRS Congress, Switzerland, 12–23.

Eddy, P. R., Smith, A. M., Hill, B. D., Peddle, D. R., Coburn, C. A., Blackshaw, R. E., 2008. Hybrid Segmentation – Artificial Neural Network Classification of High Resolution Hyperspectral Imagery for Site-Specific Herbicide Management in Agriculture 74 (10): 1249–1257.

Ehlers, M., 1991. Multisensor image fusion techniques in remote sensing. ISPRS Journal of Photogrammetry and Remote Sensing 46 (1): 19–30.

Eismann, M., 2012. Hyperspectral Remote Sensing. SPIE Press, Bellingham, Wash.

Eismann, M., Hardie, R., 2004. Application of the stochastic mixing model to hyperspectral resolution enhancement. Geoscience and Remote Sensing, IEEE Transactions on 42 (9): 1924–1933.

Eismann, M., Hardie, R., 2005. Hyperspectral resolution enhancement using high-resolution multispectral imagery with arbitrary response functions. Geoscience and Remote Sensing, IEEE Transactions on 43 (3): 455–465.

Eismann, M. T., 2004. Resolution enhancement of hyperspectral imagery using maximum a posteriori estimation with a stochastic mixing model. Ph.D. thesis, Univ. Daton, Department of Electrical and Computer Engineering, Dayton, OH.

Elad, M., 2010. Sparse and Redundant Representations. Springer New York, New York, NY.

Eldar, Y. C., Kutyniok, G., 2012. Compressed Sensing. Cambridge University Press.

Eldar, Y. C., Mishali, M., 2009. Robust Recovery of Signals From a Structured Union of Subspaces. IEEE Transactions on Information Theory 55 (11): 5302–5316.

Eldar, Y. C., Rauhut, H., 2010. Average Case Analysis of Multichannel Sparse Recovery Using Convex Relaxation. IEEE Transactions on Information Theory 56 (1): 505–519.

Ellis, J. M., Davis, H. H., Zamudio, J. A., 2001. Exploring for onshore oil seep. ResearchGate 99 (37): 49–+.

Fan, C., 2014. Spectral Analysis of Water Reflectance for Hyperspectral Remote Sensing of Water Quailty in Estuarine Water. Journal of Geoscience and Environment Protection 02 (02): 19.

Févotte, C., Dobigeon, N., 2015. Nonlinear Hyperspectral Unmixing With Robust Nonnegative Matrix Factorization. IEEE Transactions on Image Processing 24 (12): 4810–4819.

Figueiredo, M. A. T., Nowak, R. D., 2003. An EM algorithm for wavelet-based image restoration. IEEE Transactions on Image Processing 12 (8): 906–916.

Fornasier, M., 2010. Theoretical Foundations and Numerical Methods for Sparse Recovery. Walter de Gruyter.

Fornasier, M., Rauhut, H., 2008. Recovery Algorithms for Vector-Valued Data with Joint Sparsity Constraints. SIAM Journal on Numerical Analysis 46 (2): 577–613.

Foucart, S., Rauhut, H., 2013. A Mathematical Introduction to Compressive Sensing. Applied and Numerical Harmonic Analysis. Springer New York, New York, NY.

Frank, M., Pan, Z., Raber, B., Lenart, C., 2010. Vegetation management of utility corridors using high-resolution hyperspectral imaging and LiDAR. In: 2010 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, 1–4.

Galli, A., Malinverni, S., 2003. Different processing techniques to analyze high resolution hyperspectral data of urban areas. In: 2003 2nd GRSS/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas, 23–32.

Gao, C.-Z., Dinh, P. M., Reinhard, P.-G., Suraud, E., 2015. On the analysis of photo-electron spectra. arXiv:1505.04982 [physics] .

Gao, L., Du, Q., Zhang, B., Yang, W., Wu, Y., 2013. A Comparative Study on Linear Regression-Based Noise Estimation for Hyperspectral Imagery. IEEE Journal of Selected Topics in Applied Earth

Observations and Remote Sensing 6 (2): 488–498.

Gao, L.-R., Zhang, B., Zhang, X., Zhang, W.-J., Tong, Q.-X., 2008. A New Operational Method for Estimating Noise in Hyperspectral Images. IEEE Geoscience and Remote Sensing Letters 5 (1): 83–87.

Garguet-Duport, B., Girel, J., Chassery, J.-M., Pautou, G., 1996. The Use of Multiresolution Analysis and Wavelets Transform for Merging SPOT Panchromatic and Multispectral Image Data. Photogrammetric Engineering & Remote Sensing 62 (9): 1057–1066.

Garzelli, A., Nencini, F., 2009. Hypercomplex Quality Assessment of Multi/Hyperspectral Images. IEEE Geoscience and Remote Sensing Letters 6 (4): 662–665.

GDAL, 2017. GDAL - Geospatial Data Abstraction Library. www.gdal.org.

Ghamisi, P., Souza, R., Benediktsson, J. A., Rittner, L., Lotufo, R., Zhu, X. X., 2016. Hyperspectral Data Classification Using Extended Extinction Profiles. IEEE Geoscience and Remote Sensing Letters 13 (11): 1641–1645.

Gillespie, A. R., Kahle, A. B., Walker, R. E., 1987. Color enhancement of highly correlated images. II. Channel ratio and "chromaticity" transformation techniques. Remote Sensing of Environment 22 (3): 343 – 365.

Gomez, R. B., Jazaeri, A., Kafatos, M., 2001. Wavelet-based hyperspectral and multispectral image fusion. Vol. 4383, 36–42.

Gonzalez, R. C., Woods, R. E., 2006. Digital Image Processing (3rd Edition). Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

Gonzalez-Dugo, V., Hernandez, P., Solis, I., Zarco-Tejada, P. J., 2015. Using High-Resolution Hyperspectral and Thermal Airborne Imagery to Assess Physiological Condition in the Context of Wheat Phenotyping. Remote Sensing 7 (10): 13586–13605.

Goodenough, D. G., Chen, H., Gordon, P., Niemann, K. O., Quinn, G., 2012. Forest applications with hyperspectral imaging. In: 2012 IEEE International Geoscience and Remote Sensing Symposium, 7309–7312.

Goovaerts, P., Jacquez, G. M., Marcus, A., 2005. Geostatistical and local cluster analysis of high resolution hyperspectral imagery for detection of anomalies. Remote Sensing of Environment 95 (3): 351–367.

Gosnell, M. E., Anwer, A. G., Cassano, J. C., Sue, C. M., Goldys, E. M., 2016. Functional hyperspectral imaging captures subtle details of cell metabolism in olfactory neurosphere cells, disease-specific models of neurodegenerative disorders. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research 1863 (1): 56–63.

Govender, M., Chetty, K., Bulcock, H., 2007. A review of hyperspectral remote sensing and its application in vegetation and water resource studies. Water SA 33 (2): 145–152.

Gowen, A. A., Marini, F., Esquerre, C., O'Donnell, C., Downey, G., Burger, J., 2011. Time series hyperspectral chemical imaging data: Challenges, solutions and applications. Analytica Chimica Acta 705 (1–2): 272–282.

Green, A. A., Berman, M., Switzer, P., Craig, M. D., 1988. A transformation for ordering multispectral data in terms of image quality with implications for noise removal. IEEE Transactions on Geoscience and Remote Sensing 26 (1): 65–74.

Gribonval, R., Rauhut, H., Schnass, K., Vandergheynst, P., 2008. Atoms of All Channels, Unite! Average Case Analysis of Multi-Channel Sparse Recovery Using Greedy Algorithms. Journal of Fourier Analysis and Applications 14 (5-6): 655–687.

Grohnfeldt, C., Burns, T. M., Zhu, X. X., 2015. Dictionary Learning Strategies for Sparse Representation Based Hyperspectral Image Enhancement. In: Proceedings of Whispers 2015, IEEE Xplore, Tokyo, Japan, 1–4.

Grohnfeldt, C., Zhu, X. X., 2015a. Splitting the Hyperspectral-Multispectral Image Fusion Problem into Weighted Pan-sharpening Problems - The Spectral Grouping Concept. In: Proceedings of Whispers 2015, IEEE Xplore, Tokyo, Japan, 1–4.

Grohnfeldt, C., Zhu, X. X., 2015b. Towards a combined sparse representation and unmixing based hybrid hyperspectral resolution enhancement method. In: 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2872–2875.

Grohnfeldt, C., Zhu, X. X., Bamler, R., 2013. Jointly sparse fusion of hyperspectral and multispectral imagery. In: 2013 IEEE International Geoscience and Remote Sensing Symposium - IGARSS, 4090–4093.

Grohnfeldt, C., Zhu, X. X., Bamler, R., 2014. The J-SparseFI-HM Hyperspectral Resolution Enhancement Method - Now Fully Automated. In: Proceedings of the IEEE 6th Workshop on Hyperspectral Image and Signal Processing, Evolotion in Remote Sensing, IEEE Xplore, Lausanne, Switzerland, 1–4.

Gross, H. N., Schott, J. R., 1998. Application of Spectral Mixture Analysis and Image Fusion Techniques for Image Sharpening. Remote Sensing of Environment 63 (2): 85–94.

Guanter, L., Kaufmann, H., Segl, K., Foerster, S., Rogass, C., Chabrillat, S., Kuester, T., Hollstein, A., Rossner, G., Chlebek, C., Straif, C., Fischer, S., Schrader, S., Storch, T., Heiden, U., Mueller, A., Bachmann, M., Mühle, H., Müller, R., Habermeyer, M., Ohndorf, A., Hill, J., Buddenbaum, H., Hostert, P., van der Linden, S., Leitão, P. J., Rabe, A., Doerffer, R., Krasemann, H., Xi, H., Mauser, W., Hank, T., Locherer, M., Rast, M., Staenz, K., Sang, B., 2015. The EnMAP Spaceborne Imaging Spectroscopy Mission for Earth Observation. Remote Sensing 7 (7): 8830–8857.

Gurram, P., Kwon, H., Davidson, C., 2016. Coalition Game Theory-Based Feature Subspace Selection for Hyperspectral Classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 9 (6): 2354–2364.

Gursoy, O., Birdal, A. C., Ozyonar, F., Kasaka, E., 2015. Determining and Monitoring the Water Quality of Kizilirmak River of Turkey: First Results. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 7: 1469–1474.

Haertel, V. F., Shimabukuro, Y. E., 2005. Spectral linear mixing model in low spatial resolution image data. IEEE Transactions on Geoscience and Remote Sensing 43 (11): 2555–2562.

Hardie, R. C., Eismann, M. T., Wilson, G. L., 2004. MAP estimation for hyperspectral image resolution enhancement using an auxiliary sensor. IEEE Transactions on Image Processing 13 (9): 1174–1184.

Hardin, P., Hardin, A., 2013. Hyperspectral Remote Sensing of Urban Areas. Geography Compass 7 (1): 7–21.

Heldens, W., 2010. Use of airborne hyperspectral data and height information to support urban micro-climate characterisation. thesis, Bayerischen Julius-Maximilians-Universität Würzburg.

Hermosilla, T., Ruiz, L. A., Recio, J. A., Estornell, J., 2011. Evaluation of Automatic Building Detection Approaches Combining High Resolution Images and LiDAR Data. Remote Sensing 3 (12): 1188–1210.

Heylen, R., Gader, P., 2014. Nonlinear Spectral Unmixing With a Linear Mixture of Intimate Mixtures Model. IEEE Geoscience and Remote Sensing Letters 11 (7): 1195–1199.

Heylen, R., Scheunders, P., 2016. A Multilinear Mixing Model for Nonlinear Spectral Unmixing. IEEE Transactions on Geoscience and Remote Sensing 54 (1): 240–251.

Hotelling, H., 1933. Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology 24 (6): 417–441.

Huang, X., Liu, X., Zhang, L., 2014. A Multichannel Gray Level Co-Occurrence Matrix for Multi/Hyperspectral Image Texture Representation. Remote Sensing 6 (9): 8424–8445.

Huang, Y. C., Li, P., Zhang, L., Zhong, Y., 2009. Urban land cover mapping by spatial-spectral feature analysis of high resolution hyperspectral data with decision directed acyclic graph SVM. In: 2009 Joint Urban Remote Sensing Event, 1–7.

Jawak, S. D., Luis, A. J., 2013. A Comprehensive Evaluation of PAN-Sharpening Algorithms Coupled with Resampling Methods for Image Synthesis of Very High Resolution Remotely Sensed Satellite Data. Advances in Remote Sensing 02 (04): 332.

Jensen, J. R., 2015. Introductory Digital Image Processing: A Remote Sensing Perspective. Pearson Education.

Jiang, C., Zhang, H., Shen, H., Zhang, L., 2012. A Practical Compressed Sensing-Based Pan-Sharpening Method. IEEE Geoscience and Remote Sensing Letters 9 (4): 629–633.

Jiang, C., Zhang, H., Shen, H., Zhang, L., 2014. Two-Step Sparse Coding for the Pan-Sharpening of Remote Sensing Images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7 (5): 1792–1805.

Joo Kim, S., Deng, F., Brown, M. S., 2011. Visual enhancement of old documents with hyperspectral imaging. Pattern Recognition 44 (7): 1461–1469.

Karhunen, K., 1947. Über lineare Methoden in der Wahrscheinlichkeitsrechnung. Universitat Helsinki.

Keshava, N., Mustard, J. F., 2002. Spectral unmixing. IEEE Signal Processing Magazine 19 (1): 44–57.

Khan, M. M., Alparone, L., Chanussot, J., 2009. Pansharpening Quality Assessment Using the Modulation Transfer Functions of Instruments. IEEE Transactions on Geoscience and Remote Sensing 47 (11): 3880–3891.

Köhler, C. H., 2016. Airborne Imaging Spectrometer HySpex. Journal of large-scale research facilities JLSRF 2 (0): 1–6.

Krauß, T., d'Angelo, P., Schneider, M., Gstaiger, V., 2013. The fully automatic processing system CATENA at DLR. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XL-1-W1: 177–183.

Kruse, F. A., Boardman, J. W., Huntington, J. F., 2003. Comparison of airborne hyperspectral data and EO-1 Hyperion for mineral mapping. IEEE Transactions on Geoscience and Remote Sensing 41 (6): 1388–1400.

Kruse, F. A., Lefkoff, A. B., Boardman, J. W., Heidebrecht, K. B., Shapiro, A. T., Barloon, P. J., Goetz, A. F. H., 1993. The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data. Remote Sensing of Environment 44 (2): 145–163.

Laben, C. A., Brower, B. V., 2000. Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening. Patent: U.S. Classification 382/276, 382/278, 382/254, 382/299; International Classification G06T5/10; Cooperative Classification G06T5/50, G06T3/4061, G06T5/003, G06T2207/10036; European Classification G06T5/10.

Lanaras, C., 2015. Personal Website. https://www1.ethz.ch/igp/photogrammetry/people/lanaras.

Lanaras, C., Baltsavias, E., Schindler, K., 2015. Hyperspectral Super-Resolution by Coupled Spectral Unmixing. In: 2015 IEEE International Conference on Computer Vision (ICCV), IEEE, 3586–3594.

Lanaras, C., Baltsavias, E., Schindler, K., 2016. Estimation of Relative Sensor Characteristics For Hyperspectral Super-Resolution. IEEE Xplore, Los Angeles, USA.

Landgrebe, D. A., 2005. Signal Theory Methods in Multispectral Remote Sensing. John Wiley & Sons, google-Books-ID: 4Kcfg1od3twC.

Lee, J. B., Woodyatt, A. S., Berman, M., 1990. Enhancement of high spectral resolution remote-sensing data by a noise-adjusted principal components transform. IEEE Transactions on Geoscience and Remote Sensing 28 (3): 295–304.

Legleiter, C. J., 2003. Spectrally Driven Classification of High Spatial Resolution, Hyperspectral Imagery: A Tool for Mapping In-Stream Habitat. Environmental Management 32 (3): 399–411.

Lein, J. K., 2012. Environmental Sensing. Springer New York, New York, NY.

Lenhard, K., Baumgartner, A., Gege, P., Nevas, S., Nowy, S., Sperling, A., 2015a. Impact of Improved Calibration of a NEO HySpex VNIR-1600 Sensor on Remote Sensing of Water Depth. IEEE Transactions on Geoscience and Remote Sensing 53 (11): 6085–6098.

Lenhard, K., Baumgartner, A., Schwarzmaier, T., 2015b. Independent Laboratory Characterization of NEO HySpex Imaging Spectrometers VNIR-1600 and SWIR-320m-e. IEEE Transactions on Geoscience and Remote Sensing 53 (4): 1828–1841.

Li, C., Yang, X., Chu, B., Lu, W., Pang, L., 2010. A new image fusion quality assessment method based on Contourlet and SSIM. In: 2010 3rd International Conference on Computer Science and Information Technology, Vol. 5, 246–249.

Li, S., Yang, B., 2011. A New Pan-Sharpening Method Using a Compressed Sensing Technique. IEEE Transactions on Geoscience and Remote Sensing 49 (2): 738–746.

Li, W., Du, Q., Zhang, F., Hu, W., 2016. Hyperspectral Image Classification by Fusing Collaborative and Sparse Representations. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 9 (9): 4178–4187.

Li, W., Tramel, E. W., Prasad, S., Fowler, J. E., 2014. Nearest Regularized Subspace for Hyperspectral Classification. IEEE Transactions on Geoscience and Remote Sensing 52 (1): 477–489.

Li, X., Gao, Z., Bai, L., Huang, Y., 2012. Potential of high resolution RapidEye data for sparse vegetation fraction mapping in arid regions. In: 2012 IEEE International Geoscience and Remote Sensing Symposium, 420–423.

Liaw, A., Wiener, M., 2001. Classification and Regression by RandomForest. ResearchGate 23 (3).

Licciardi, G. A., Veganzones, M. A., Simoes, M., Bioucas-Dias, J. M., Chanussot, J., 2014. Super-resolution of hyperspectral images using local spectral unmixing. In: IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS 2014), Lausanne, Switzerland, n/c.

Lillesand, T., Kiefer, R. W., Chipman, J., 2014. Remote Sensing and Image Interpretation. John Wiley & Sons.

Lillo-Saavedra, M., Gonzalo, C., 2007. Multispectral images fusion by a joint multidirectional and multiresolution representation. International Journal of Remote Sensing 28 (18): 4065–4079.

Liu, J., Wang, H., Qin, W., 2011. A new fusion image quality assessment based on edge and structure similarity. In: 2011 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems, 112–115.

Liu, J. G., 2000. Smoothing Filter-based Intensity Modulation: A spectral preserve image fusion technique for improving spatial details. International Journal of Remote Sensing 21 (18): 3461–3472.

Liu, X., Tanaka, M., Okutomi, M., 2013. Single-Image Noise Level Estimation for Blind Denoising. IEEE Transactions on Image Processing 22 (12): 5226–5237.

Locherer, M., 2014. Capacity of the hyperspectral satellite mission EnMAP for the multiseasonal monitoring of biophysical and biochemical land surface parameters in agriculture by transferring an analysis method for airborne image spectroscopy to the spaceborne scale. Phd thesis, Ludwig-Maximilians-Universität München.

Lombardo, V., Harris, A. J. L., Calvari, S., Buongiorno, M. F., 2009. Spatial variations in lava flow field thermal structure and effusion rate derived from very high spatial resolution hyperspectral (MIVIS) data. Journal of Geophysical Research: Solid Earth 114 (B2): B02208.

Loncan, L., de Almeida, L. B., Bioucas-Dias, J. M., Briottet, X., Chanussot, J., Dobigeon, N., Fabre, S., Liao, W., Licciardi, G. A., Simoes, M., Tourneret, J.-Y., Veganzones, M. A., Vivone, G., Wei, Q., Yokoya, N., 2015. Hyperspectral Pansharpening: A Review. IEEE Geoscience and Remote Sensing Magazine 3 (3): 27–46.

López-López, M., Calderón, R., González-Dugo, V., Zarco-Tejada, P. J., Fereres, E., 2016. Early Detection and Quantification of Almond Red Leaf Blotch Using High-Resolution Hyperspectral and Thermal Imagery. Remote Sensing 8 (4): 276.

LRZ, 2017. Leibniz Supercomputing Centre. https://www.lrz.de.

Maglione, P., 2016. Very High Resolution Optical Satellites: An Overview of the Most Commonly used. American Journal of Applied Sciences 13 (1): 91–99.

Mairal, J., Bach, F., Ponce, J., Sapiro, G., 2009. Online Dictionary Learning for Sparse Coding. In: Proceedings of the 26th Annual International Conference on Machine Learning, ICML '09, ACM, New York, NY, USA, 689–696.

Makarau, A., Palubinskas, G., Reinartz, P., 2012. Selection of numerical measures for pan-sharpening

assessment. In: 2012 IEEE International Geoscience and Remote Sensing Symposium, 2264–2267.
Mallat, S. G., Zhang, Z., 1993. Matching pursuits with time-frequency dictionaries. IEEE Transactions on Signal Processing 41 (12): 3397–3415.
Manolakis, D., Siracusa, C., Shaw, G., 2001. Hyperspectral subpixel target detection using the linear mixing model. IEEE Transactions on Geoscience and Remote Sensing 39 (7): 1392–1409.
Marino, C. M., Panigada, C., Busetto, L., 2001. Airborne hyperspectral remote sensing applications in urban areas: Asbestos concrete sheeting identification and mapping. In: IEEE/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas (Cat. No.01EX482), 212–216.
Martin, M. E., Plourde, L. C., Ollinger, S. V., Smith, M. L., McNeil, B. E., 2008. A generalizable method for remote sensing of canopy nitrogen across a wide range of forest ecosystems. Remote Sensing of Environment 112 (9): 3511–3519.
Meikle, H., 2008. Modern Radar Systems, 2nd Edition. Artech House Publishers, Boston.
Melgani, F., Bruzzone, L., 2004. Classification of hyperspectral remote sensing images with support vector machines. IEEE_J_GRS 42 (8): 1778–1790.
Middleton, E. M., Ungar, S. G., Mandl, D. J., Ong, L., Frye, S. W., Campbell, P. E., Landis, D. R., Young, J. P., Pollack, N. H., 2013. The Earth Observing One (EO-1) Satellite Mission: Over a Decade in Space. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 6 (2): 243–256.
Mielke, C., Rogass, C., Boesche, N., Segl, K., Altenberger, U., 2016. EnGeoMAP 2.0—Automated Hyperspectral Mineral Identification for the German EnMAP Space Mission. Remote Sensing 8 (2): 127.
Mitchell, P. A., 1995. Hyperspectral digital imagery collection experiment (HYDICE). Vol. 2587, 70–95.
Mitri, G. H., Gitas, I. Z., 2010. Mapping Postfire Vegetation Recovery Using EO-1 Hyperion Imagery. IEEE Transactions on Geoscience and Remote Sensing 48 (3): 1613–1618.
Mitri, G. H., Gitas, I. Z., 2013. Mapping post-fire forest regeneration and vegetation recovery using a combination of very high spatial resolution and hyperspectral satellite imagery. International Journal of Applied Earth Observation and Geoinformation 20: 60–66.
Monteiro, S., Minekawa, Y., Kosugi, Y., Akazawa, T., Oda, K., 2006. Prediction of Sweetness and Nitrogen Content in Soybean Crops from High Resolution Hyperspectral Imagery. In: 2006 IEEE International Symposium on Geoscience and Remote Sensing, 2263–2266.
Müller, R., Avbelj, J., Carmona, E., Gerasch, B., Graham, L., Günther, B., Heiden, U., Kerr, G., Knodt, U., Krutz, D., Krawczyk, H., Makarau, A., Miller, R., Perkins, R., Walter, I., 2016. The New Hyperspectral Sensor DESIS on the Multi-Payload Platform MUSES Installed on the ISS. Prague, Czech Republic.
Müller, R., Bachmann, M., Makasy, C., de Miguel, A., Müller, A., Neumann, A., Palubinskas, G., Richter, R., Schneider, M., Storch, T., Walzel, T., Kaufmann, H., Guanter, L., Segl, K., Heege, T., Kiselev, V., 2010. The processing chain and Cal/Val operations of the future hyperspectral satellite mission EnMAP. In: 2010 IEEE Aerospace Conference, 1–9.
Murphy, R. J., Monteiro, S. T., Schneider, S., 2012. Evaluating Classification Techniques for Mapping Vertical Geology Using Field-Based Hyperspectral Sensors. IEEE Transactions on Geoscience and Remote Sensing 50 (8): 3066–3080.
NASA, 1992. EOS Data and Information System (EOSDIS) - NASA-TM-107922.
Nascimento, J. M. P., Dias, J. M. B., 2005a. Signal Subspace Identification in Hyperspectral Linear Mixtures. In: Pattern Recognition and Image Analysis, Springer, Berlin, Heidelberg, 207–214.
Nascimento, J. M. P., Dias, J. M. B., 2005b. Vertex component analysis: A fast algorithm to unmix hyperspectral data. IEEE Transactions on Geoscience and Remote Sensing 43 (4): 898–910.
Nasrabadi, N. M., 2014. Hyperspectral Target Detection : An Overview of Current and Future Challenges. IEEE Signal Processing Magazine 31 (1): 34–44.
Natarajan, B., 1995. Sparse Approximate Solutions to Linear Systems. SIAM Journal on Computing

24 (2): 227–234.

Nencini, F., Garzelli, A., Baronti, S., Alparone, L., 2007. Remote sensing image fusion using the curvelet transform. Information Fusion 8 (2): 143–156.

Nicodemus, F. E., 1965. Directional Reflectance and Emissivity of an Opaque Surface. Applied Optics 4 (7): 767.

Nikolakopoulos, K. G., 2009. Spatial resolution enhancement of Hyperion hyperspectral data. In: 2009 First Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, 1–4.

Noviello, M., Ciminale, M., De Pasquale, V., 2013. Combined application of pansharpening and enhancement methods to improve archaeological cropmark visibility and identification in QuickBird imagery: Two case studies from Apulia, Southern Italy. Journal of Archaeological Science 40 (10): 3604–3613.

Numata, I., Cochrane, M. A., Galvão, L. S., 2010. Analyzing the Impacts of Frequency and Severity of Forest Fire on the Recovery of Disturbed Forest using Landsat Time Series and EO-1 Hyperion in the Southern Brazilian Amazon. Earth Interactions 15 (13): 1–17.

Nunez, J., Otazu, X., Fors, O., Prades, A., Pala, V., Arbiol, R., 1999. Multiresolution-based image fusion with additive wavelet decomposition. IEEE Transactions on Geoscience and Remote Sensing 37 (3): 1204–1211.

Obata, K., Tsuchida, S., Nagatani, I., Yamamoto, H., Kouyama, T., Yamada, Y., Yamaguchi, Y., Ishii, J., 2016. An overview of ISS HISUI hyperspectral imager radiometric calibration. In: 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 1924–1927.

OpenRemoteSensing, 2015. Source Codes. `http://www.openremotesensing.net/index.php/codes/11-pansharpening`.

Otazu, X., Gonzalez-Audicana, M., Fors, O., Nunez, J., 2005. Introduction of sensor spectral response into image fusion methods. Application to wavelet-based methods. IEEE Transactions on Geoscience and Remote Sensing 43 (10): 2376–2385.

Pahlevan, N., Wei, J., Schaaf, C. B., Schott, J. R., 2014. Evaluating radiometric sensitivity of Landsat 8 over coastal/inland waters. In: 2014 IEEE Geoscience and Remote Sensing Symposium, 1393–1396.

Pearson, K., 1901. On lines and planes of closest fit to systems of points in space. Philosophical Magazine Series 6 2 (11): 559–572.

Peng, Z., Yan, M., Yin, W., 2013. Parallel and distributed sparse optimization. In: 2013 Asilomar Conference on Signals, Systems and Computers, 659–646.

Peshkun, A., Oct. 20-22, 2014. Resurs-P capabilities and standard products. In: From Imagery to Map Digital Photogrammetric Technologies, China.

Peter, S., 2016. Algorithms for Robust and Fast Sparse Recovery. Dissertation, Technische Universität München, Munich, Germany.

Pirondini, F., Al Marri, S., 2014. The DUBAISAT-2/DEIMOS-2 constellation: Public-private cooperation between Emirates and Spain. Vol. 9241, 92411D–92411D–9.

Pirondini, F., López, J., González, E., González, J. A., 2014. DEIMOS-2: Cost-effective, very-high resolution multispectral imagery. Vol. 9241, 92411C–92411C–14.

Pistonesi, S., Martinez, J., Ojeda, S. M., Vallejos, R., 2015. A Novel Quality Image Fusion Assessment Based on Maximum Codispersion. In: Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. Springer International Publishing, 383–390.

Pohl, C., 2016. Multisensor image fusion guidelines in remote sensing. IOP Conference Series: Earth and Environmental Science 34 (1): 012026.

Pohl, C., Genderen, J. L. V., 1998. Review article Multisensor image fusion in remote sensing: Concepts, methods and applications. International Journal of Remote Sensing 19 (5): 823–854.

Pohl, C., van Genderen, J., 2015. Structuring contemporary remote sensing image fusion. International Journal of Image and Data Fusion 6 (1): 3–21.

Polo, G., Labruna, M. B., Ferreira, F., 2015. Satellite Hyperspectral Imagery to Support Tick-Borne Infectious Diseases Surveillance. PLOS ONE 10 (11): e0143736.

Popescu, S. C., Zhao, K., Neuenschwander, A., Lin, C., 2011. Satellite lidar vs. small footprint airborne lidar: Comparing the accuracy of aboveground biomass estimates and forest structure metrics at footprint level. Remote Sensing of Environment 115 (11): 2786–2797.

Proctor, C., He, Y., 2015. Workflow for building a hyperspectral UAV: Challenges and oppertunities. In: ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XL-1-W4, Copernicus GmbH, 415–419.

Pyatykh, S., Hesser, J., Zheng, L., 2013. Image Noise Level Estimation by Principal Component Analysis. IEEE Transactions on Image Processing 22 (2): 687–699.

Qayyum, A., Malik, A. S., Saad, M. N. B. M., 2015. Evaluation of digital elevation model using rational polynomial coefficient based on HR Pleiades satellite stereo imagery. In: 2015 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), 32–37.

Rahmani, S., Strait, M., Merkurjev, D., Moeller, M., Wittman, T., 2010. An Adaptive IHS Pan-Sharpening Method. IEEE Geoscience and Remote Sensing Letters 7 (4): 746–750.

Ranchin, T., Wald, L., 2000. Fusion of high spatial and spectral resolution images: The arsis concept and its implementation. Photogrammetric Engineering and Remote Sensing 66 (1): 49–61.

Raol, J. R., 2015. Data Fusion Mathematics. CRC Press.

Rasti, B., Ulfarsson, M. O., Sveinsson, J. R., 2015a. Hyperspectral Subspace Identification Using SURE. IEEE Geoscience and Remote Sensing Letters 12 (12): 2481–2485.

Rasti, B., Ulfarsson, M. O., Sveinsson, J. R., 2015b. Hyperspectral Subspace Identification Using SURE. IEEE Geoscience and Remote Sensing Letters 12 (12): 2481–2485.

Regniers, O., Bombrun, L., Guyon, D., Samalens, J. C., Germain, C., 2015. Wavelet-Based Texture Features for the Classification of Age Classes in a Maritime Pine Forest. IEEE Geoscience and Remote Sensing Letters 12 (3): 621–625.

Richards, J. A., 2013. Remote Sensing Digital Image Analysis: An Introduction. Springer Science & Business Media.

Richter, R., Schlapfer, D., Muller, A., 2011. Operational Atmospheric Correction for Imaging Spectrometers Accounting for the Smile Effect. IEEE Transactions on Geoscience and Remote Sensing 49 (5): 1772–1780.

Ristin, M., Guillaumin, M., Gall, J., Gool, L. V., 2016. Incremental Learning of Random Forests for Large-Scale Image Classification. IEEE Transactions on Pattern Analysis and Machine Intelligence 38 (3): 490–503.

Robinson, G. D., Gross, H. N., Schott, J. R., 2000. Evaluation of Two Applications of Spectral Mixing Models to Image Fusion. Remote Sensing of Environment 71 (3): 272–281.

Rodriguez, J. J., Kuncheva, L. I., 2006. Rotation forest: A new classifier ensemble method. IEEE_J_PAMI 28 (10): 1619–1630.

Roger, R. E., 1996. Principal Components transform with simple, automatic noise adjustment. International Journal of Remote Sensing 17 (14): 2719–2727.

Roger, R. E., Arnold, J. F., 1996. Reliably estimating the noise in AVIRIS hyperspectral images. International Journal of Remote Sensing 17 (10): 1951–1962.

Sarvotham, S., Baron, D., Wakin, M., Duarte, M. F., Baraniuk, R. G., 2005. Distributed Compressed Sensing of Jointly Sparse Signals. 1537–1541.

Schneider, S., Murphy, R. J., Melkumyan, A., 2014. Evaluating the performance of a new classifier – the GP-OAD: A comparison with existing methods for classifying rock type and mineralogy from hyperspectral imagery. ISPRS Journal of Photogrammetry and Remote Sensing 98: 145–156.

Schölkopf, B., Burges, C. J. C., Smola, A. J. (Hrsg.), 1999. Advances in Kernel Methods: Support Vector Learning. MIT Press, Cambridge, MA, USA.

Schowengerdt, R. A., 2006. Remote Sensing, Third Edition: Models and Methods for Image Processing, 3rd Edition. Academic Press, Burlington, MA.

Segl, K., Guanter, L., Gascon, F., Kuester, T., Rogass, C., Mielke, C., 2015. S2eteS: An End-to-End Modeling Tool for the Simulation of Sentinel-2 Image Products. IEEE Transactions on Geoscience and Remote Sensing 53 (10): 5560–5571.

Segl, K., Guanter, L., Rogass, C., Kuester, T., Roessner, S., Kaufmann, H., Sang, B., Mogulsky, V., Hofer, S., 2012. EeteS – The EnMAP End-to-End Simulation Tool. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 5 (2): 522–530.

Selva, M., Aiazzi, B., Butera, F., Chiarantini, L., Baronti, S., 2015. Hyper-Sharpening: A First Approach on SIM-GA Data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 8 (6): 3008–3024.

Shafri, H. Z. M., Taherzadeh, E., 2012. Hyperspectral Remote Sensing of Urban Areas: An Overview of Techniques and Applications. ResearchGate 4 (11): 1557–1565.

Shah, V. P., Younan, N. H., King, R. L., 2008. An Efficient Pan-Sharpening Method via a Combined Adaptive PCA Approach and Contourlets. IEEE Transactions on Geoscience and Remote Sensing 46 (5): 1323–1335.

Shaw, G. A., Burke, H.-h. K., 2003. Spectral Imaging for Remote Sensing. ResearchGate 14 (1).

Shobiga, R., Selvakumar, J., 2015. Survey on properties and accuracy assessment of climate changes using hyperspectral imaging. In: 2015 Online International Conference on Green Engineering and Technologies (IC-GET), 1–6.

Shu, Y., Tang, H., Li, J., Mao, T., He, S., Gong, A., Chen, Y., Du, H., 2015. Object-Based Unsupervised Classification of VHR Panchromatic Satellite Images by Combining the HDP and IBP on Multiple Scenes. IEEE Transactions on Geoscience and Remote Sensing 53 (11): 6148–6162.

Simões, M., Bioucas-Dias, J., Almeida, L. B., Chanussot, J., 2015. A Convex Formulation for Hyperspectral Image Superresolution via Subspace-Based Regularization. IEEE Transactions on Geoscience and Remote Sensing 53 (6): 3373–3388.

Slater, P. N., 1980. Remote Sensing: Optics and Optical Systems. Addison-Wesley Publishing Company.

Solomon, C., Breckon, T., 2011. Fundamentals of Digital Image Processing: A Practical Approach with Examples in Matlab. John Wiley & Sons.

Spoto, F., Sy, O., Laberinti, P., Martimort, P., Fernandez, V., Colin, O., Hoersch, B., Meygret, A., 2012. Overview Of Sentinel-2. In: 2012 IEEE International Geoscience and Remote Sensing Symposium, 1707–1710.

Starck, J. L., Donoho, D. L., Candès, E. J., 2003. Astronomical image representation by the curvelet transform. Astronomy & Astrophysics 398 (2): 785–800.

Stefano, P., Angelo, P., Simone, P., Filomena, R., Federico, S., Tiziana, S., Umberto, A., Vincenzo, C., Acito, N., Marco, D., Stefania, M., Giovanni, C., Raffaele, C., Roberto, D. B., Giovanni, L., Cristina, A., 2013. The PRISMA hyperspectral mission: Science activities and opportunities for agriculture and land monitoring. In: 2013 IEEE International Geoscience and Remote Sensing Symposium - IGARSS, 4558–4561.

Stewart, G., 1993. On the Early History of the Singular Value Decomposition. SIAM Review 35 (4): 551–566.

Stoll, E., Konstanski, H., Anderson, C., Douglass, K., Oxfort, M., 2012. The RapidEye constellation and its data products. In: 2012 IEEE Aerospace Conference, 1–9.

Strait, M., Rahmani, S., Merkurjev, D., 2008. Evaluation of Pan-Sharpening Methods. Technical report, University of California, Los Angeles (UCLA), Los Angeles, CA, U.S.A.

Sun, L., Wu, Z., Liu, J., Xiao, L., Wei, Z., 2015. Supervised Spectral #x2013;Spatial Hyperspectral Image Classification With Weighted Markov Random Fields. IEEE Transactions on Geoscience and Remote Sensing 53 (3): 1490–1503.

SuperMUC, 2017. LRZ – SuperMUC Petascale System. https://www.lrz.de/services/compute/supermuc/systemdescription/.

Svetnik, V., Liaw, A., Tong, C., Culberson, J. C., Sheridan, R. P., Feuston, B. P., 2003. Random Forest:

A Classification and Regression Tool for Compound Classification and QSAR Modeling. ResearchGate 43 (6): 1947–58.

Tanii, J., Kashimura, O., Ito, Y., Iwasaki, A., 2016. Flight model performances of HISUI hyperspectral sensor onboard ISS (International Space Station). Vol. 10000, 100000A–100000A–7.

Teke, M., Deveci, H. S., Haliloğlu, O., Gürbüz, S. Z., Sakarya, U., 2013. A short survey of hyperspectral remote sensing applications in agriculture. In: 2013 6th International Conference on Recent Advances in Space Technologies (RAST), 171–176.

Thomas, C., Ranchin, T., Wald, L., Chanussot, J., 2008. Synthesis of Multispectral Images to High Spatial Resolution: A Critical Review of Fusion Methods Based on Remote Sensing Physics. IEEE Transactions on Geoscience and Remote Sensing 46 (5): 1301–1312.

Tibshirani, R., 1996. Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society. Series B (Methodological) 58 (1): 267–288.

Tinel, C., Grizonnet, M., Fontannaz, D., de Boissezon, H., Giros, A., 2012. ORFEO, The Pleiades Accompaniment Program and its Users Thematic Commissioning. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXIX-B3: 569–572.

Ting-ting, Z., Fei, L., 2012. Application of hyperspectral remote sensing in mineral identification and mapping. In: Proceedings of 2012 2nd International Conference on Computer Science and Network Technology, 103–106.

Top500, 2016. List of the most powerful commercially available computer systems. https://www.top500.org/project/top500_description/.

Torti, E., Acquistapace, M., Danese, G., Leporati, F., Plaza, A., 2014. Real-Time Identification of Hyperspectral Subspaces. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7 (6): 2680–2687.

Tropp, J. A., 2006. Algorithms for simultaneous sparse approximation. Part II: Convex relaxation. Signal Processing 86 (3): 589–602.

Tropp, J. A., Gilbert, A. C., 2007. Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit. IEEE Transactions on Information Theory 53 (12): 4655–4666.

Tropp, J. A., Gilbert, A. C., Strauss, M. J., 2006. Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit. Signal Processing 86 (3): 572–588.

Tsai, F., Chen, W. W., 2008. Striping Noise Detection and Correction of Remote Sensing Images. IEEE Transactions on Geoscience and Remote Sensing 46 (12): 4122–4131.

Tu, T.-M., Huang, P. S., Hung, C.-L., Chang, C.-P., 2004. A fast intensity-hue-saturation fusion technique with spectral adjustment for IKONOS imagery. IEEE Geoscience and Remote Sensing Letters 1 (4): 309–312.

Tu, T.-M., Su, S.-C., Shyu, H.-C., Huang, P. S., 2001. A new look at IHS-like image fusion methods. Information Fusion 2 (3): 177–186.

tuxfamily, 2017. The Eigen library - C++ template library for linear algebra. http://eigen.tuxfamily.org.

van den Berg, E., Friedlander, M., 2008. Probing the Pareto Frontier for Basis Pursuit Solutions. SIAM Journal on Scientific Computing 31 (2): 890–912.

van der Meer, F. D., de Jong, S. M. (Hrsg.), 2013. Imaging Spectrometry: Basic Principles and Prospective Applications, 2001st Edition. Springer, Dordrecht; Boston.

Vandenabeele, P., Donais, M. K., 2016. Mobile Spectroscopic Instrumentation in Archaeometry Research. Applied Spectroscopy 70 (1): 27–41.

Vane, G., Green, R. O., Chrien, T. G., Enmark, H. T., Hansen, E. G., Porter, W. M., 1993. The airborne visible/infrared imaging spectrometer (AVIRIS). Remote Sensing of Environment 44 (2): 127–143.

Vapnik, V., 2000. The Nature of Statistical Learning Theory. Springer-Verlag New York, Inc., New York, NY, USA.

Vierling, L. A., Vierling, K. T., Adam, P., Hudak, A. T., 2013. Using Satellite and Airborne LiDAR to Model Woodpecker Habitat Occupancy at the Landscape Scale. PLOS ONE 8 (12): 1–13.

Vivone, G., Alparone, L., Chanussot, J., Mura, M. D., Garzelli, A., Licciardi, G. A., Restaino, R., Wald, L., 2015. A Critical Comparison Among Pansharpening Algorithms. IEEE Transactions on Geoscience and Remote Sensing 53 (5): 2565–2586.

Vreys, K., Iordache, M.-D., Bomans, B., Meuleman, K., 2016. Data acquisition with the APEX hyperspectral sensor. Miscellanea Geographica 20 (1): 5–10.

Wald, L., 2000. Quality of high resolution synthesised images: Is there a simple criterion? In: Proceedings of the 3rd International Conference on Fusion of Earth Data, SEE/URISCA, Sophia Antipolis, France, 99–103.

Walsh, S. J., McCleary, A. L., Mena, C. F., Shao, Y., Tuttle, J. P., González, A., Atkinson, R., 2008. QuickBird and Hyperion data analysis of an invasive plant species in the Galapagos Islands of Ecuador: Implications for control and land use management. Remote Sensing of Environment 112 (5): 1927–1941.

Wang, T., Yan, G., Ren, H., Mu, X., 2010. Improved Methods for Spectral Calibration of On-Orbit Imaging Spectrometers. IEEE Transactions on Geoscience and Remote Sensing 48 (11): 3924–3931.

Wang, Z., Bovik, A. C., 2002. A universal image quality index. IEEE Signal Processing Letters 9 (3): 81–84.

Wang, Z., Ziou, D., Armenakis, C., Li, D., Li, Q., 2005. A comparative analysis of image fusion methods. IEEE Transactions on Geoscience and Remote Sensing 43 (6): 1391–1402.

Waske, B., van der Linden, S., Benediktsson, J. A., Rabe, A., Hostert, P., 2010. Sensitivity of Support Vector Machines to Random Feature Selection in Classification of Hyperspectral Data. IEEE Transactions on Geoscience and Remote Sensing 48 (7): 2880–2889.

Wei, Q., 2015. Personal Website. `http://wei.perso.enseeiht.fr/publications.html`.

Wei, Q., Bioucas-Dias, J., Dobigeon, N., Tourneret, J. Y., 2015a. Hyperspectral and Multispectral Image Fusion Based on a Sparse Representation. IEEE Transactions on Geoscience and Remote Sensing 53 (7): 3658–3668.

Wei, Q., Dobigeon, N., Tourneret, J. Y., 2015b. Bayesian Fusion of Multi-Band Images. IEEE Journal of Selected Topics in Signal Processing 9 (6): 1117–1127.

Wei, Q., Dobigeon, N., Tourneret, J. Y., 2015c. Fast Fusion of Multi-Band Images Based on Solving a Sylvester Equation. IEEE Transactions on Image Processing 24 (11): 4109–4121.

Wei, X., Jianguo, W., Yinzhu, C., 2014. Fusion of Remote Sensing Image with Compressed Sensing Based on Wavelet Sparse Basis. In: 2014 Sixth International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), 287–289.

White, J. C., Gómez, C., Wulder, M. A., Coops, N. C., 2010. Characterizing temperate forest structural and spectral diversity with Hyperion EO-1 data. Remote Sensing of Environment 114 (7): 1576–1589.

Wiedermann, G., Gockel, W., Winkler, S., Rieber, J. M., Kraft, B., Reggio, D., 2014. The Sentinel-2 Satellite Attitude Control System – Challenges and Solutions. In: ResearchGate.

Wu, C., Du, B., Zhang, L., 2013. A Subspace-Based Change Detection Method for Hyperspectral Images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 6 (2): 815–830.

Wu, C., Wang, L., Niu, Z., Gao, S., Wu, M., 2010. Nondestructive estimation of canopy chlorophyll content using Hyperion and Landsat/TM images. International Journal of Remote Sensing 31 (8): 2159–2167.

Wu, J.-L., Ho, C.-R., Huang, C.-C., Srivastav, A. L., Tzeng, J.-H., Lin, Y.-T., 2014. Hyperspectral Sensing for Turbid Water Quality Monitoring in Freshwater Rivers: Empirical Relationship between Reflectance and Turbidity and Total Solids. Sensors (Basel, Switzerland) 14 (12): 22670–22688.

Wycoff, E., Chan, T. H., Jia, K., Ma, W. K., Ma, Y., 2013. A non-negative sparse promoting algorithm for high resolution hyperspectral imaging. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 1409–1413.

Xia, J., Du, P., He, X., Chanussot, J., 2014. Hyperspectral remote sensing image classification based

on rotation forest. IEEE_J_GRSL 11 (1): 239–243.
Xiong, W., Shao, Y., Shen, W., Xiao, R., Fu, Z., Shi, Y., 2014. One method for HJ-1-A HSI and CCD data fusion. IOP Conference Series: Earth and Environmental Science 17 (1): 012226.
Yamamoto, H., Kouyama, T., Obata, K., Tsuchida, S., 2015. Assessment of HISUI radiometric performance using vicarious calibration and cross-calibration. In: 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2805–2808.
Yang, S., Shi, Z., 2016. Hyperspectral Image Target Detection Improvement Based on Total Variation. IEEE Transactions on Image Processing 25 (5): 2249–2258.
Yang, X.-H., Jiao, L.-C., 2008. Fusion Algorithm for Remote Sensing Images Based on Nonsubsampled Contourlet Transform. Acta Automatica Sinica 34 (3): 274–281.
Yokoya, N., 2012. Source Code – CNMF. http://naotoyokoya.com/Download.html.
Yokoya, N., Grohnfeldt, C., Chanussot, J., 2017. Hyperspectral and Multispectral Data Fusion: A comparative review of the recent literature. IEEE Geoscience and Remote Sensing Magazine 5 (2): 29–56.
Yokoya, N., Iwasaki, A., 2016. Airborne hyperspectral data over Chikusei. Tech. rep., Space Application Laboratory, the University of Tokyo, Japan.
Yokoya, N., Mayumi, N., Iwasaki, A., 2013. Cross-Calibration for Data Fusion of EO-1/Hyperion and Terra/ASTER. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 6 (2): 419–426.
Yokoya, N., Yairi, T., Iwasaki, A., 2011a. Coupled non-negative matrix factorization (CNMF) for hyperspectral and multispectral data fusion: Application to pasture classification. In: 2011 IEEE International Geoscience and Remote Sensing Symposium, 1779–1782.
Yokoya, N., Yairi, T., Iwasaki, A., 2011b. Hyperspectral, multispectral, and panchromatic data fusion based on coupled non-negative matrix factorization. In: 2011 3rd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 1–4.
Yokoya, N., Yairi, T., Iwasaki, A., 2012. Coupled Nonnegative Matrix Factorization and Unmixing for Hyperspectral and Multispectral Data Fusion. IEEE Transactions on Geoscience and Remote Sensing 50 (2): 528–537.
Yuen, P. W., Richardson, M., 2010. An introduction to hyperspectral imaging and its application for security, surveillance and target acquisition. The Imaging Science Journal 58 (5): 241–253.
Yuhendra, Alimuddin, I., Sumantyo, J. T. S., Kuze, H., 2012. Assessment of pan-sharpening methods applied to image fusion of remotely sensed multi-band data. International Journal of Applied Earth Observation and Geoinformation 18: 165–175.
Yuhendra, J. T., Sumantyo, H., Kuze, H., 2011. Performance analyzing of high resolution pan-sharpening techniques: Increasing image quality for classification using supervised kernel Support Vector Machine. Research Journal of Information Technology 3 (1): 12–23.
Zelentsov, V., Sokolov, B., Grigorieva, O., Mochalov, V., Potryasaev, S., Shumeiko, V., 2016. "resurs-P" satellite hyperspectral data: Preliminary evaluation of information capacities. In: European Space Agency, (Special Publication) ESA SP, Vol. SP-740, European Space Agency.
Zeng, Y., Zhang, J., Wang, G., Li, Y., 2011. RapidEye Satellite Image Quality Analysis and Solutions for Its True Color Composition. In: 2011 International Workshop on Multi-Platform/Multi-Sensor Remote Sensing and Mapping (M2RSM), 1–4.
Zhang, H., j Song, H., c Yu, B., 2011. Application of Hyper Spectral Remote Sensing for Urban Forestry Monitoring in Natural Disaster Zones. In: 2011 International Conference on Computer and Management (CAMAN), 1–4.
Zhang, L., Zhu, X., Zhang, L., Du, B., 2016. Multidomain Subspace Classification for Hyperspectral Images. IEEE Transactions on Geoscience and Remote Sensing 54 (10): 6138–6150.
Zhang, Y., 1999. A new merging method and its spectral and spatial effects. International Journal of Remote Sensing 20 (10): 2003–2014.
Zhang, Y., He, M., 2007. Multi-spectral and hyperspectral image fusion using 3-D wavelet transform.

Journal of Electronics (China) 24 (2): 218–224.
Zhang, Y., He, M., 2009. 3D wavelet transform and its application in multispectral and hyperspectral image fusion. In: 4th IEEE Conference on Industrial Electronics and Applications, 2009. ICIEA 2009, 3643–3647.
Zhao, X., Xiao, Z., Kang, Q., Li, Q., Fang, L., 2010. Overview of the Fourier Transform Hyperspectral Imager (HSI) boarded on HJ-1A satellite. In: Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International, 4272–4274.
Zhong, Z., Fan, B., Duan, J., Wang, L., Ding, K., Xiang, S., Pan, C., 2015. Discriminant Tensor Spectral #x2013;Spatial Feature Extraction for Hyperspectral Image Classification. IEEE Geoscience and Remote Sensing Letters 12 (5): 1028–1032.
Zhou, J., Civco, D. L., Silander, J. A., 1998. A wavelet transform method to merge Landsat TM and SPOT panchromatic data. International Journal of Remote Sensing 19 (4): 743–757.
Zhu, X. X., Bamler, R., 2010. Tomographic SAR Inversion by -Norm Regularization – The Compressive Sensing Approach. IEEE Transactions on Geoscience and Remote Sensing 48 (10): 3839–3846.
Zhu, X. X., Bamler, R., 2012. Super-Resolution Power and Robustness of Compressive Sensing for Spectral Estimation With Application to Spaceborne Tomographic SAR. IEEE Transactions on Geoscience and Remote Sensing 50 (1): 247–258.
Zhu, X. X., Bamler, R., 2013. A Sparse Image Fusion Algorithm With Application to Pan-Sharpening. IEEE Transactions on Geoscience and Remote Sensing 51 (5): 2827–2836.
Zhu, X. X., Grohnfeldt, C., Bamler, R., 2013a. Collaborative sparse reconstruction for pan-sharpening. In: Geoscience and Remote Sensing Symposium (IGARSS), 2013 IEEE International, 868–871.
Zhu, X. X., Grohnfeldt, C., Bamler, R., 2013b. Collabrative sparse image fusion with application to pan-sharpening. In: 2013 18th International Conference on Digital Signal Processing (DSP), 1–6.
Zhu, X. X., Grohnfeldt, C., Bamler, R., 2016. Exploiting Joint Sparsity for Pansharpening: The J-SparseFI Algorithm. IEEE Transactions on Geoscience and Remote Sensing 54 (5): 2664–2681.
Zhukov, B., Oertel, D., Lanzl, F., Reinhackel, G., 1999. Unmixing-based multisensor multiresolution image fusion. IEEE Transactions on Geoscience and Remote Sensing 37 (3): 1212–1226.

Acknowledgement

This dissertation is the result of many people's support who I would like to thank sincerely.

First, I would like to express my gratitude to Munich Aerospace e.V. and its staff for financially supporting and administrating my work in the scope of the SparsEO project. Furthermore, I would like to thank the Gauss Centre for Supercomputing e.V. for providing computing time on the SuperMUC at the Leibniz Supercomputing Centre.

None of the achievements that are summarized in this thesis would have been possible, if it hadn't been for the permanent financial and scientific support, patience and inspiring discussions with Prof. Xiaoxiang Zhu, whom I sincerely thank for the fruitful time and everything I learned while pursuing my Ph.D. under her excellent supervision. Her comprehensive knowledge in countless fields, sophistication, unconstrained and interdisciplinary way of thinking, talent for picking up new theories and constant desire to innovate has been inspiring me from the moment I joined her team at the DLR.

I am equally grateful to my supervisor, Prof. Richard Bamler, whose ingenuity has so often impressed, inspired and motivated me and whose supervision, time and support I highly appreciate.

Moreover, I would like to express my sincere gratitude to Prof. Naoto Yokoya, who has so brilliantly and patiently supported and supervised me in many ways. I wish there would have been more time for collaboration during our common time at the DLR. He has been a mentor and friend from the moment we met in Australia in 2013. I hope, that our friendship and common interests will make our paths cross often in the future.

I have been lucky that I was able to work together and make friends with so many brilliant, inspiring and cordial people, both as a member of the SiPEO Team at the TUM and as a member of the SAR Signal Processing Department at the DLR. Although every one deserves to be named, I would like to explicitly mention Dr. Yuanyuan Wang, Prof. Muhammad Shahzad and Mr. Sina Montazeri, whom I have shared many moments of excitement, frustration and everything in between with.

I am tremendously thankful to my fellow student, roommate, friend and project partner, Dr. Steffen Peter, who has been a friend and supportive far beyond the scope of the SparsEO project. I am equally grateful to my friend and former student, Mr. Tristan Burns, whose positive attitude, excitement for science and ingenuity have often sparked creativity in me.

Last but not least, I dedicate this graduation to my dear Negin and to my family. My gratefulness and love for you is beyond words.